Abstract

Homeopathic medicine remains widely used in Britain and around the world. The evidence on which homeopaths base decisions about which homeopathic medicines to prescribe comes largely from homeopathic pathogenetic trials (HPTs).Such trials aim to produce pathogenetic effects in healthy participants by administering repeated doses of a homeopathic medicine. A previous systematic review of the quality of all such trials performed between 1945 and 1995 found that they were generally of poor quality.

This review looked at all HPTs' performed up to 2009 and found that a small percentage of the total (3%) of such trials were of high quality. 15 met rigorous inclusion criteria and had methods which were likely to keep bias to a minimum. Trials used the number and pattern of symptoms occurring in participants as outcomes measures; these were recorded by participants in structured and unstructured diaries and questionnaires.

An analysis of the outcomes of such trials failed to find evidence that Homeopathic medicines produce pathogenetic effects in healthy adults any different to those which occur in participants taking identical placebo medicines.

The trials which met the inclusion criteria were generally of high methodological quality. However, none of them used procedures to screen or select participants who were likely to be sensitive to the particular medicine used in the trial. It is recommended that future trials adopt such a screening process.

Table of contents

Acknowledgements

I would like to acknowledge the support which I have received from my supervisors, Morag Heirs and Catherine Hewitt. Their great patience and detailed guidance have been crucial in the process of turning ideas into appropriate words. I am not a fan of Oscar style paeans of praise but offer a simple thank-you to all others who have helped me along the way.

Background

1.1 Homeopathy- the contemporary context

homeopathy, developed in Germany some 200 years ago (Lockie 1989), remains a controversial but highly popular form of health care intervention worldwide. The World Health Organisation estimate that 500 million people worldwide use homeopathy (WHO 2005) and in the UK it is estimated that 15% of the population use and trust it (TGI 2008). It has also always been a part of the National Health Service, and there are currently four homeopathic NHS hospitals receiving 55,000 referrals a year and 400 GP's practising homeopathy and treating 200,000 patients per year with homeopathic medicine within primary care.(British homeopathic Association 2009).

It is important that an intervention which is so widely used, is evaluated in a robust and scientific manner, using accepted and appropriate methods. There are valid debates about the most appropriate methods for evaluating the types of complex interventions which take place in complementary therapies such as homeopathy. (Weatherley Jones et al 2004). However, it is widely acknowledged that the most robust research designs for evaluating all health care interventions, and those least susceptible to various forms of bias are randomised controlled trials and systematic reviews of such trials. (Torgerson and Torgerson 2008, Bowling and Ebrahim 2005)

It has been suggested that extraordinary claims require extraordinary standards of evidence (Truzzi 1998).homeopathy, which involves serial dilution of substances to ultra molecular levels, appears implausible and extraordinary. Whilst research methods may be need to be adapted to suit the way in which the therapy is practiced, homeopathy should certainly be subjected to at least the same standards of evidence as any other treatment or intervention.

The mechanism of action of homeopathy remains controversial and unknown (Bellavite and Signorini 1995) and much scepticism exists as to whether homeopathic medicines can have any effect beyond placebo. However, a range of research methods have been brought to bear on the study of homeopathy in recent

years. Basic laboratory research (see e.g. Wagner et al 1988 Bellavite et al 1991, Jonas et al 2001, Fimiani 2000), experimental research on animals (e.g. De Gerlache and Lans 1991, Bildet et al 1990, Conforti et al 1993, Toper et al 1990) and numerous human clinical trials have all taken place. The Hombrex database of basic research currently contains details of over 1300 scientific experiments on the effects of homeopathic preparations on biological systems. (Carstens Stiftung 2009). In terms of clinical trials on human subjects, the most recent meta-analysis (Shang et al 2005) assessed 110 placebo controlled trials.

1.2 The Importance of HPTs (Homeopathic Pathogenetic Trials)

The major principle of homeopathic medicine is the idea of similars that like cures like. What this suggests in practice is that any substance which can cause symptoms in a healthy person may cure the same symptoms in a person who is ill.

Given that this is the major theoretical idea in homeopathy, it is crucial to test the idea experimentally and to determine whether substances can cause pathogenetic effects in healthy persons and the nature of any such effects.

Samuel Hahnemann, who first developed the system of homeopathic medicine, developed a methodology for formally testing the properties of substances by testing them on groups of healthy people. The information from these tests then suggested the ways in which the medicines might be used in treating patients. Hahnemann named these trials ‘prufungs’, a term which became translated as “provings”. More recently such tests have become known as homeopathic Pathogenetic Trials or HPTs.(Dantas et al 2007)

The knowledge base on which homeopathy rests results from a triangulation of information: from HPTs, from clinical practice and, where relevant, from toxicology if the medicine is question is based on a substance which is toxic in its crude state. (Belon 1995) Of these, HPTs remain one of the foundation stones of the practice of homeopathic medicine. (Dominici et al 2007 Koster et al 1998)

Thus HPTs have a central role in homeopathy, both theoretically and in practice. Most clinical decisions taken by a homeopathic practitioner about which treatment to

use are based, to a greater or lesser degree, on information gleaned from such HPTs. It is therefore very important to survey the knowledge base which makes up the HPT literature and to assess the quality and reliability of such information. The clinical practice of homeopathy can only be effective, and will only demonstrate efficacy in research, if the information base for practice is accurate.

1.3 The Methodology of HPTs

Hahnemann developed a methodology which was quite rigorous for the time. He first developed the idea for these trials in 1790 in correspondence with the local authorities regarding the urgent need for medical reform (Belon 1995). “You *should choose medicines for the symptoms they produce in healthy individual’s bodies as witnessed by repeated observation*”. He went on to test over 100 different medicines individually in quasi –experimental studies. (De Marque 1987). To minimise bias he recommended that only conscientious and trustworthy volunteers who were healthy at the time of the trial should take part. Also, that medicines should be tested singly and in pure form with close supervision of subjects throughout the duration of the trial. Hahnemann recommended quite strict rules to control for what he believed were important confounding variables. His aim was to ensure a constant steady state in which any changes caused by the medicine would be easier to detect and he therefore recommended that the consumption of tea, coffee, alcohol, medicines, herbs and spices and any foods which might have medicinal effects should be avoided. Hahnemann was convinced that medicine needed to progress via clinical experiment rather than by theoretical debate. (Dean 2004). Here he was clearly an early champion of what later became standard procedure for all drug trials.

As Dantas et al (2006) have noted, HPTs have been methodologically innovative. The first blinded placebo controlled “proving” took place in a Naval hospital in St Petersburg in 1834 and a year later a double blind trial was carried out in Nuremberg which even attempted random allocation (Dean 2004).One of the earliest recorded multi- centre double blind clinical trials in the history of medicine was an HPT of Belladonna in 1906. (Bellows 1906)

Despite these aspects of methodological rigour and innovation in some HPTs, subsequent analysis has shown that Hahnemann's guidelines have a number of weaknesses many of which may have led to overestimation of pathogenetic effects for various reasons. (Dantas 1996). These include: the absence of a control group as standard; the use of well known friends as volunteers; the absence of blinding; close daily supervision and recording of trivial and pre-existing symptoms (leading to naturally occurring and existing symptoms being recorded as pathogenetic and the occurrence of Hawthorne and recall effects); the sudden prohibition of all medicinal drugs and foodstuffs (leading to the effects of abstinence being included as pathogenetic symptoms); and vague definitions of healthy (leading to the inclusion of symptoms of existing disease being included as pathogenetic).

However, during the twentieth century the limited number of HPT's of homeopathy continued to follow the guidance of Hahnemann. It was only with the publication of 'The Dynamics and Methodology of homeopathic Provings' by Jeremy Sherr in 1994 that any significant new guidance in relation to HPTs emerged (Sherr1994).This was the first publication which was focussed entirely on HPTs and has been very influential for the design of many HPTs which have been conducted since that date. Sherr described a clear methodology for the conduct of HPTs. However, his methodology does not address a number of the weaknesses identified by Dantas. Notably, he neglects the important role of placebo control as a comparator for the active intervention. Sherr identifies three major benefits of placebo in drug trials, and admits that the first of these, distinguishing the effects of actual treatment from the effects of the test process, is relevant. However, he dismisses the other benefits of placebo, suggesting that distinguishing drug effects from fluctuations in disease that occur with time and other external factors does not apply as provings are invariably made on healthy volunteers. This is an unwarranted assumption, since the definition and measurement of health in participants in HPTs varies widely, and ,as will be shown later, the most healthy of individuals do experience fluctuations in symptoms over the course of a trial, whatever efforts are made to avoid other influences.

Subsequently various sets of guidelines were developed. One such came from the European Committee for Homeopathy (ECH 2004). Their drug proving group developed very detailed protocols following a series of annual international symposia

which had begun in 1992, and they first published guidance in 1997. In 2004 they updated their guidelines and based them on the structure and contents of the Guidelines for Good Clinical Practice of the International Conference on Harmonisation (ICH). Initially a tripartite agreement between the EU, the Unites States and Japan, the ICH guidelines have been widely adopted worldwide and are aimed at producing uniform standards in relation to the design and conduct of medical trials.(European Medicines Agency 2006). In attempting to bring guidelines for HPTs up to the standards of the ICH, the ECH made a wide range of recommendations, including: minimum qualifications for the principal investigator; the definition and reporting of adverse events and serious adverse events; the need for explicit protocols; assessment of safety; and the summary of the known risks and benefits to participants in the trial.

Guidelines published by Riley in 1997 also emphasised the need to address ethical issues, to describe the origin and preparation of the homeopathic medicine in detail, to have clear inclusion and exclusion criteria and to have very clear methods and guidelines for symptom selection. (Riley 1997)

Walach (1997) suggested that different methodologies would be needed for different purposes, though they might be combined. He recommended that trials aimed at determining that homeopathic substances can produce symptoms different from placebo should adopt placebo controlled double blind designs and to avoid carry over effects, should be of parallel group design. In contrast trials designed to improve practical prescribing in homeopathy and particularly those aimed at finding new medicines would need to follow the traditional qualitative methodology which had been used for 200 years for HPTs.

Herscu (2002) gathered together a range of historical and contemporary writings on HPTs and developed his own guidelines. He outlined a ten step process which included most of the steps suggested by other guidance, but also included a recommendation to retest the provers who are most sensitive to the medicine in the initial trial

Table 1 offers a comparison of the key published guidance in relation to HPTs

Table 1 Comparison of key items in published guidance on the conduct of HPTs

Author	Hahne-mann 1821	Sherr 1994	Dantas 1996	Riley 1997	Walach 1997	Herscu 2002	ECH 2004
Use of placebo	NO	10-20%	YES	NO	YES	12.5%	YES
Parallel group design	NO	NO	YES	NO	YES	NO	NO
Randomisation and allocation concealment procedures	NO	NO	YES	NO	YES	NO	YES
Blinding	NO	NO	YES	NO	YES	YES	YES
Ethics approval and informed consent	NO	NO	NO	YES	YES	YES	YES
Validated definition of healthy participant at entry	NO	NO	YES	NO	NO	NO	NO
Procedures for dealing with missing data/loss to follow up	NO	NO	NO	YES	NO	NO	YES
Procedures for screening for participants who are sensitive to	NO	NO	NO	NO	NO	YES	NO

the trial medicine							

All of these guidelines have led to many recommendations for trial design. In practice, the methodological quality of both the design and conduct of clinical and pathogenetic trials in homeopathy remains variable though and many poor quality trials have continued to be conducted (Linde 2001, Dantas et al 2006). It should be noted that in this respect the situation is no different to that which pertains in other areas of health care research.

Some HPTs have been performed which incorporate the standards which have become the norm in medical and health related trials and have addressed the main factors which are known to introduce bias and to lead to an overestimation of effects. These standards include: placebo controls, clear processes of random allocation, of allocation concealment and of blinding of participants and researchers; as well as explicit processes for dealing with drop outs and missing data .One purpose of this review was to systematically review such trials, and to determine how many HPT's were of high methodological quality.

1.4 Susceptibility to pathogenetic effects

In terms of methodology, it is crucial that HPT's are designed in ways which best facilitate the observation and measurement of any pathogenetic effects. There is a particular problem which is perhaps unique to HPTs which relates to susceptibility. It is known that people respond in a range of ways to the administration of any form of medicine. Whether healthy or sick, a person taking any form of medicine may display a range of responses. What has been called the 'total drug response' will be a result of the action of the medicine together with a range of non- specific effects and placebo effects. (Peters 1991). One of the non specific effects relates to the individual susceptibility.

All participants will react to sufficient doses of substance in a crude state, since any substance if given in sufficient quantity will exert some physiological and eventually

toxicological effects. Even water, in sufficiently large doses over a short period of time, can lead to an acute intoxication and can lead to fatal reactions such hyponatremia, which, for example, has caused deaths in several marathon runners in recent years (McClatchy 2001, Almond et al 2005)

However, when substances are successively diluted in the manner used in homeopathy they show less toxicological or direct physiological effects with increasing dilution. (refs ???????)The process involved in the preparation of homeopathic medicines involves a series of stages of both dilution and succussion (vigorous shaking of the dilution). Insoluble substances are first triturated with lactose to a point at which they become soluble; other substances are diluted from the crude state in a mixture of alcohol and water.

Dilution usually takes place at the level of one part in ten or one part in a hundred. In practice there are certain common levels of dilution which are used as medicines, and these are denoted with a system of number and letters. (see table 1.2 for details of typical dilutions and nomenclature used in Homeopathic pharmacy)

Table 1.2 Homeopathic medicines. Levels of dilution and succession of commonly used medicines.

Designation	Level of dilution	Number of stages of dilution and succussion
1x or D1	One part in ten	1
6x Or D6	One part in ten	6
6C	One part in a hundred	6
30 C	One part in a hundred	30
200C	One part in a hundred	200
1M	One part in a hundred	1000
LM1	One part in fifty thousand	1

T

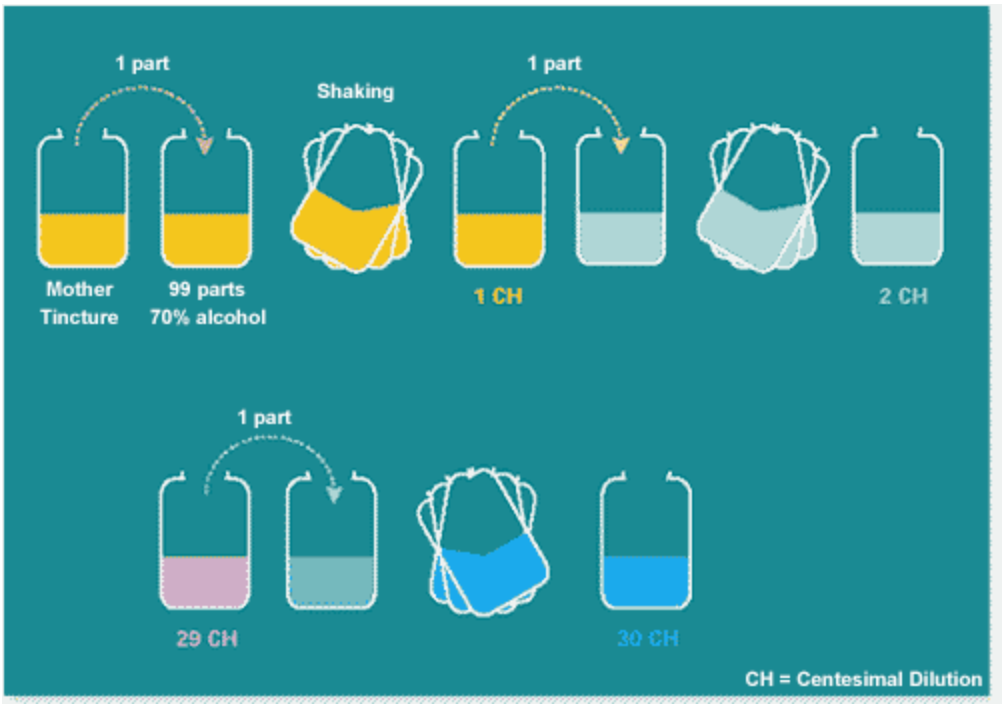


Fig 1.1 Illustration of the stages of production of a 30C strength homeopathic medicine

This process means that increasing potencies (as they are known) contain less of the original substance and it is thought that with each increase in potency less people will be sensitive or susceptible to the effects of that medicine. (Herscu 2002). Moreover, because effects will only occur in individuals with a predisposition to be affected by the substance and the constitution and predisposition of each person varies somewhat, it has been observed that certain symptoms will occur in some individuals and others in different individuals. Hahnemann suggested that "the *total picture of disease symptoms that a (homeopathic) medicine can produce approaches completion only after multiple observations have been made on suitable persons with various constitutions*". (Hahnemann 1921).

The theoretical basis of homeopathy suggests that only persons who have a particular constitutional and biological make up will be susceptible to the effects of a specific homeopathically prepared medicine and will respond to it, and each may respond differently. (Hahnemann 1921, Vithoulkas 1981). According to this model of individual susceptibility this will be true both in clinical practice and also in HPTs. (refs …..

This means that any individual symptom may be a rare event and unlikely to be detected using a standard RCT designs. In theory pooling data from different trials and using meta –analysis can be helpful (if appropriate) when events rates are very low (Deeks et al 1998), and suggests another reason for attempting a systematic review with meta-analysis of HPTs.

The issue of rare events and trial design is often discussed in relation to the detection of adverse events in drug trials. However, there is a further complication with HPT's in that there is not a single symptom or even a small number of possible pathogenetic effects. Kiene (1995) has pointed out that there are an enormous number of potential symptom outcomes in a HPT and this means that statistical detection of individual symptoms distinct from placebo and background noise is unlikely using conventional statistics.

One possible approach to this problem might be to take homeopathic medicines which have already been subject to HPTs and to widespread clinical use. Outcome measures could then focus on a limited number of key common symptoms which are known to be linked to the medicine. This approach was taken in some of the trials included in the current review.

If the purpose of an HPT is to demonstrate that there is a statistically significant difference between verum and placebo then this approach may be useful. If the purpose is to develop a complete record of pathogenetic effects which might be useful for clinical practice then this approach is inadequate. Kaptchuk (1996) has concluded that "homeopathy *still has no clear answer to the question or rare symptoms versus chance symptoms in proving*".

It is not known what proportion of participants are likely to be sensitive to the effects of a medicine during an HPT though the potency used will be have an impact on this. It is thought that with increasing "potencies" (which are increasingly diluted and succussed) homeopathic medicines will progressively target only those who are most susceptible to their effects. For this reason most HPT's use relatively low potencies. Dantas et al noted that 30C was the most commonly used, followed by 6C and 6X. The latter is the lowest potency which is generally available in homeopathic medicines.

The theory clearly suggests that only a minority will respond with pathogenetic effects to a medicine given in a homeopathic potency, though this minority may be larger with the use of lower potencies. Many HPT's in the past appear to have found that all participants displayed pathogenetic effects. (Dantas et al 2007, Herscu 2002)

As Dantas et al noted in their review this is likely to be an artefact of poor methodological quality and to reflect non - specific effects, Hawthorne effects, and the nature of the participants (usually enthusiastic students of homeopathy eager and primed to expect pathogenetic effects). In their review of many HPTs these authors noted that 84% of volunteers in verum groups displayed symptoms, and that in the trials of lower methodological quality all participants reported pathogenetic effects. They give the example of two trials of the same medicine, which are of very different methodological quality. The low quality trial (Sherr 1992) reported 5000% more pathogenetic effects per participant than the high quality one (Schroyens 1996).

Homeopathic theory suggests that treatment should always be individualised to the patient. It is possible to design trials of clinical homeopathic practice which allow for individualisation of treatment and still maintain randomisation and blinding, since the practitioner has the opportunity to interview the patient and individually select a remedy before randomisation occurs. If the practitioner achieves accuracy in selecting the medicine to which each individual patient is susceptible then it is likely to show an effect in each patient. The situation of HPTs is different. Typically participants taking part are volunteers who meet inclusion criteria and are willing. There is no procedure for determining who may be susceptible to the remedy being tested before the trial begins. But to reiterate homeopathic theory suggests that only persons who are susceptible will respond, whether they are healthy participants being exposed to repeated doses of the remedy in a pathogenetic trial or whether they are ill patients who are prescribed the remedy to treat their symptoms. In clinical practice the Homeopath will use their skill and knowledge to match the susceptibility of the patient to the most similar remedy. For HPTs it follows that only those who have some susceptibility to the medicine being tested will react and show symptoms, and it must be important to have some kind of procedure for pre selecting

volunteers who are sensitive enough to show pathogenetic effects from the particular medicine used in the trial.

Shalts has summarised this problem by noting that “one *of the reasons that proving studies consistently show no difference is that the studies do not allow for people who are not sensitive to verum. They also do not allow for the variable symptoms that arise in every participant*.” (Shalts 2002)

One of the aims of this review was to look at methods used to select and determine the sensitivity of participants before or during trials.

1.5 Placebo and nocebo

HPTs are aimed at producing pathogenetic effects in participants. Placebo effects can also be pathogenetic as well as beneficial. These are sometimes known as nocebo effects (Ernst 2001).

One of the challenges for homeopathic pathogenetic trials is to establish whether the specific signal of the medicine being tested can be distinguished from the background noise of existing symptoms, placebo effects and other non specific effects which may lead to the occurrence of new symptoms. In order to establish this it is of course necessary to have carefully controlled trials with placebo groups. Placebo effects are usually understood as therapeutic or beneficial effects occurring after the administration of some kind of placebo intervention.

Whilst the definition of a placebo intervention remains contested, it should strictly be limited to those things which are imitations of an intervention with known effects, and does not include the many non specific factors which are known to have an impact on outcomes (Kienle and Kiene 2001). Factors which are often responsible for what are sometimes labelled as placebo effects include: the natural course of disease; spontaneous, irregular fluctuations in symptoms; regression to the mean, additional treatment, giving inaccurate answers to please clinicians (the word placebo after all means to please); Hawthorne effects (the tendency of people to change their behaviour and the things that they are aware of when they are the target of special interest);and placebo interventions which actually have some active ingredient. (Kienle and Kiene 2001). Well conducted randomised controlled trials should control

for these non specific factors and mean that they are spread evenly across the intervention and placebo groups. Any differences in outcomes should then reflect a difference between placebo effects from the placebo intervention and real effects of the verum intervention.

1.6 The nature size and direction of placebo effects.

When the effects of, and responses to placebo have been examined in trials it is clear that the number of participants who demonstrate any placebo response varies, that the pattern and direction of placebo response varies and that the response is modulated by many factors, including expectation. For example participants who were informed that they would definitely receive a placebo treatment showed a different response to those who were unsure whether they would receive placebo or an active intervention (Peters 2001). In another example, medical students who were conditioned to expect sedative or stimulant effects, but who received pink or blue placebo capsules, demonstrated significant placebo effects related to the expected response (Blackwell et al 1972). Ernst suggests that patient expectation is probably one of the strongest determinants of the placebo response (Ernst 2001).

To further illustrate the strength of this phenomenon, placebo responses can include significant, measurable physiological changes such as airway reactivity (Kemeny et al 2007); neurobiological responses have been measured in a range of trials and it is clear that opiate and dopamine pathways are activated and modulated in response to expectations (de la Fuentefernández and Stoessl 2002).

Placebo effects often seem to mimic those of active interventions and dose effect relationships (Blackwell, Bloomfield and Buncher 1972), cumulative effects after repeated administration and carry over effects after cessation have all been described (Lasagna, Laties and Dohan 1950)

Most recently researchers have suggested that differences in genotypes which affect the enzymes which produce both dopamine and norepinephrine (both neurotransmitters involved in reward pathways in the brain) affect responses to placebo interventions (Leuchter 2009)

For a long period of time following a seminal paper by Beecher in 1955, there was an assumption that placebo effects occur in a fairly constant 35% of participants in studies. The original data on which Beecher based his figures has subsequently been shown to be seriously flawed and work by Richardson (1994) Ernst (2001) and others has shown that the actual figure varies between 0 % and 100% depending on the context .Nocebo effects occur in most placebo controlled clinical trials and may affect up to 40 % of those receiving a placebo intervention (Tangrea, Adrianza and Helsel 1994). It was noted earlier that HPTs have some similarities with Phase 1 drug trials. Nocebo symptoms in healthy persons are commonly reported in such trials (Sibille et al 1992). Nocebo effects take various forms and can include the induction of pain (Benedetti et al 2006). It is known that conditioning and expectation increase nocebo effects. (Johansen 2003) Several studies have found higher rates of nocebo effects in female participants. (Liccardi et al 2004, Casper et al 2001).

However, to illustrate the lack of consensus about the nature and magnitude of placebo effects. two systematic reviews by Hjobartson and Gotsche (2004) of trials which compared placebo interventions with no treatment, concluded that placebo effects are very small. They found evidence of placebo effects only for patient reported outcomes and, particularly those relating to pain.

These authors acknowledge that “our *results do not exclude the possibility that other aspects of the patient–provider interaction, or interactions between the treatment ritual and different ways of informing patients, could have clinically useful effects.”* Because participants receiving no treatment also interact with treatment providers there may be factors relating to relationships and contexts which lead to similar outcomes in both groups. However, those receiving no treatment would not have the same expectations as those receiving placebo and expectation clearly did not have the same influence on outcomes in the trials considered for the review as in the literature previously described.

It seems clear that studies and reviews report wide variations in the nature and size of placebo effects and there is not yet a clear understanding of how these operate in trials of health interventions.

1.7 Background noise

Pathogenetic effects need to be distinguished from placebo/nocebo effects. They also need to be distinguished from what might be called ‘background noise’. Although HPTs are designed to use only healthy volunteers, it is clear that all people healthy or not, experience a fluctuating level of symptoms on a day to day basis. Kaptchuk (1996) has highlighted some studies which illustrate the size and variability of this background noise.

In one trial 4800 people were told that a new medicine was being tested for side effects. Participants resided in a variety of settings, including prisons, old people’s homes and medical professional offices. Interviews and pre-treatment questionnaires were designed to look for pre treatment symptoms.

Whilst the overall mean incidence of symptoms was 10%, at least half of the symptoms occurred in 20% or more of individuals, and those in certain settings had very high levels (e.g. 28% of those in old people’s homes reported dizziness) (Green 1964). In a second study 414 people who worked or studied at a university medical centre were healthy in the sense of reporting no current illness and no medication use in the previous 3 days. Of this sample 81% had at least one symptom in the previous 72 hours based on a 25 symptom adverse event questionnaire. Kaptchuk notes that the symptom check list looked very like the tabulated results of an HPT.

It is clear then that trials of interventions can lead to placebo effects of significant magnitude and in significant numbers of participants. To show that an intervention is different from placebo it must be able to produce effects of large enough magnitude and in large enough numbers of participants to outweigh both placebo effects which may already be large, and existing background noise which may also be of large magnitude, in order to reach statistical significance.

One of the ways in which background noise can be measured and differentiated in trials is to have a baseline or observation period before any intervention (placebo or verum) begins.

As discussed earlier, another method which can help to control for background noise is to predefine symptoms which might be expected from a particular medicine and to use outcome measures related specifically to such symptoms.

It has already been noted that any percentage of people (from 0 to 100) may show effects when taking placebo, that nocebo symptoms typically occur in 40% of participants and that any sample of even a healthy population will contain many individuals, possibly up to 80% of the total, with existing symptoms. Also, that only a small percentage of participants are likely to be sensitive to a particular homeopathic medicine and to show true pathogenetic effects in HPTs.

It follows from this that the most carefully controlled trials are likely to show little or no difference between verum and placebo groups in HPTs in relation to the number of individuals showing symptoms, unless the participants have already been screened and selected for their sensitivity before being recruited and randomised. There may be a difference in the pattern of symptoms in the limited number of those who do respond but not in the number of those responding.

It is notable that this factor seems to have been frequently overlooked in the design of HPTs. A lot of attention has been paid to the selection of participants in terms of their health status but not sufficient to issues of sensitivity. Some authors, notably Vithoulkas (1980) and Herscu (2002) have highlighted the issue and recommended trial designs for HPTs which incorporate a stage of selection for sensitivity/susceptibility.

There should be a process, perhaps involving several stages, to determine which participants may show symptoms. Otherwise the problem exists that the percentage of individuals who are susceptible to the substance being tested may be small and may not exceed the percentage of responders that may be deemed to be due to chance.

One way in which Homeopaths have dealt with this problem is to suggest that verification of whether a symptom is a true symptom of the remedy or is background noise or a placebo effect, will come from clinical practice rather than from HPTs. If a medicine is used in an ill person and a symptom which was listed in the HPT disappears , then that is confirmation of the homeopathic principle of like cures like and verification of that symptom. However, as Kapthcuk notes, whilst this approach makes sense in many ways, it does not help the methodology of HPT's and only postpones the problem (Kaptchuk 1996)

1.8 The nature of pathogenetic effects in HPTs.

homeopathic theory suggests that pathogenetic effects which occur in HPTs will show characteristic patterns, and that these patterns are likely to be different from placebo effects and from the type of general fluctuation in day to day symptoms which healthy people may experience. Each medicine is thought to produce a characteristic or signature pattern of symptoms.

It may be possible to predict these symptoms in advance if previous HPTs have been carried out or if information about the particular medicine is available from toxicology or from clinical practice. Also, the response of the human organism to homeopathic medicines is thought to involve a departure from homeostasis followed by a dynamic response from the organism in an attempt to re-establish homeostasis (Bellavite 2001, Vithoulkas 1991).Such a response may lead to intense new symptoms, the return of old symptoms, the occurrence of symptoms in particular parts of the body and symptoms with particular modalities. In contrast it is known that the most frequently reported nocebo effects in trials are headache, drowsiness, tiredness, dizziness, nausea, pain and insomnia (Rosenwels, Brochier and Zipfel 1993)

HPT's should use outcome measures which address and capture such patterns of symptoms rather than simply recording the numbers of symptoms in verum and control groups.

The forgoing discussion of susceptibility to pathogenetic effects, the nature size and direction of placebo effects and issues of background noise, illustrates some of the

challenges which face those designing HPTs. It remains the case that placebo controlled trials with blinding form one of the best trial methods for reducing overall bias. Trials which do not use comparator groups as controls to assess the level of placebo symptoms are likely to lead to an overestimation of pathogenetic effects from trials. On the other hand if quantitative statistical analysis of the different responses between intervention and comparator groups is relied upon exclusively then an underestimation of pathogenetic effects is highly likely, particularly without a staged approach which allows for the selection and recruitment of sensitive volunteers for the main trial.

1.9 Adverse events

In a systematic review of the adverse effects of homeopathy, Dantas and Rampes (2000) found that adverse effects of homeopathic medicines in controlled clinical trials were greater than placebo, but these were generally minor, transient and comparable. Of course, in HPT's the purpose of the trial is to produce pathogenetic effects and it is important to consider when a pathogenetic effect may be considered an adverse event.

Pathogenetic effects which are significant may lead to participants withdrawing or being withdrawn from trials and this may bias and complicate the collection and analysis of data. HPT's in which data is not collected from those who withdraw because of adverse events related to pathogenetic effects, are likely to underestimate overall pathogenetic effects. Under reporting of adverse events is common in medical literature and due to widely held beliefs that homeopathic medicines are safe, the phenomenon is thought to be greater in homeopathic literature (Dantas and Rampes 2007).Some of the guidelines for HPTs which were referred to earlier do explicitly state the steps that should be taken in relation to dealing with and recording adverse events. See, for example the ECH guidelines (ECH 1997). Again, it seems to be the case that many trials which have taken place have not followed such guidance. Dantas et al (2007) in their review of HPTs noted *"an extremely low value of withdrawals due to severe adverse effects".* They further noted that 85% of reports did not mention how adverse effects were managed. The RedHot guidelines on the reporting of trials of homeopathy (Dean et al 2007) note

that the issue of aggravation of symptoms following the administration of homeopathic medicines should be added to category 19 (reporting of adverse events) of the CONSORT statement on the reporting of trials.

The reporting of adverse events was one of the aspects of trials which this review considered, as part of an assessment of the overall reporting and management of loss to follow up.

1.10 Systematic reviews

As noted earlier systematic reviews of well conducted randomised controlled trials are considered to be one of the best sources of evidence in health care. This is because they help to answer focussed questions in a manner which is as free from bias as possible. They help to ensure that a complete assessment of existing trial literature is made using a transparent and replicable process, in a standardised way. The rigorous methodology of systematic reviews was developed to ensure that problems of subjectivity, selectivity and timeliness which occurred in more traditional narrative reviews were avoided (Cullinan 2005).

Whilst there are several systematic reviews and Meta –analyses of homeopathy in general (Kleijnen 1991, Linde 1997, Shang et al 2005) and of homeopathic treatment for specific conditions (e.g. McCarney et al 2004, Heirs and Dean 2009) there are no Systematic reviews of HPTs in the standard databases (DARE, Cochrane collaboration)

There exists a systematic review of the quality of Provings (Dantas et al 2007). entitled “A *systematic review of the quality of homeopathic pathogenetic trials published from 1945 to 1995”.*

This review has the merits of being very comprehensive and covering papers published in six languages. An international team of 13 authors searched published reports in English, German, Spanish, French, Portugese and Dutch from 1945 to 1995. With a focus on issues of quality, the authors assessed trials using an 87 item checklist which included items relating to: the description of medicines, volunteers,

ethical aspects, blinding, randomisation, and use of placebo, adverse effects, assessments, presentation of data, and the number of claimed findings. Their approach to quality was to score each of four key components: randomisation, blinding, inclusion and exclusion criteria, and criteria for selection of effects. Overall scores were then ranked in four methodological classes. The authors then discussed differences in findings and claimed results in relation to methodological quality. They noted that trials with low methodological rigour reported more effects per participant. They also noted an increase in the quality of trials towards the end of the review period, since the introduction of modern reporting guidelines such as Consort (Altman 1996).

In their paper they suggest that "*We hope ...our study ... will stimulate a close monitoring and comparison of methodological quality of HPTs done after 1995*." They also note that "*the methodological quality showed a trend to improvement in the later decades, there was a positive and significant correlation between methodological classes and decades (*r_s*=0,218; P=0,006)*". They also note that, for example, randomisation was first mentioned in 1961 but of the 15 reports which mention it, nine were in the final decade of the review period.

This suggests that the amount of primary research in conducting HPTs may have increased since 1995 and that also the quality of those may have improved , given that there has been much more discussion and, as noted earlier, a number of published guidelines in relation to quality and methodology in HPT's , as in health research generally since the 1990's . As Riley puts it "*Methodological rigor in homeopathic drug proving has probably been lacking historically; however, GCP guidelines, ethic commissions, formal protocols, and clinical trial registries are a recent invention and one should not automatically discount the historical homeopathic proving literature for not having used tools that were not available".* (Riley 1997)

The Dantas et al review was a review of the quality of trials and was not designed to synthesise data relating to outcomes. They used exploratory meta- analysis to assess some data relating to quality, and did not make any attempt at quantification

of results or meta-analysis. This means that there is no systematic review of HPT's in existence which also has a meta-analysis of quantitative outcomes.
Also, the authors had chosen a cut off date of 1995 for inclusion of studies, and it took the international team several years to complete their findings and bring them to publication (in 2007). Given that many of the guidelines discussed in 1.3, relating to the conduct and methodology of HPTs, were published in 1997 or later, there is a clear need for an updated review of HPT's. At the outset of this review a question remained as to how much HPT's had changed, improved and taken on board such guidelines since 1995 and this is a question which the review aimed to answer.

For these reasons a decision was taken to conduct a systematic review of HPTs in order to update the work started by Dantas et al but also to revisit the whole of the literature, using a different set of inclusion criteria with the aim of offering a greater degree of data synthesis, and meta –analysis relating to outcomes of HPTs, if at all possible.

Chapter Two - Methodology

2.1 Formulating a research question

There is broad agreement that systematic reviews of literature relating to health care interventions are best framed using the following components: 1.Population. 2 Intervention 3. Comparator 4.Outcome. 5 Study Design. (CRD 2009)

After an initial period of reading and development of ideas as described in Chapter one and after consultation with supervisors the following question and components for a review were formulated.

Question: To determine whether the effects of homeopathic substances (ultra molecular dilutions) on human subjects differ from the effects of placebo in homeopathic pathogenetic trials. (HPTs)

The purpose of the review was to assess studies of the effects of homeopathic substances on healthy participants in homeopathic pathogenetic trials, which have some similarities with phase one drug trials. It is standard procedure to include only healthy participants in pathogenetic trials (Wieland 1997) though few authors have operationalised the definition of healthy or have used validated measures to assess the state of health of participants in HPTs. The question of whether homeopathic substances have effects when used clinically as a treatment is a different one and the question needed to be framed in such a way as to make this clear.

2.2 Selection of studies: Assessment of relevance for inclusion in the review. inclusion and exclusion criteria.

2.2.1 *Population*

Inclusion criteria

The review considered studies involving adult participants, male and female, aged 18 years or over.

2.2.2 *Interventions*

Inclusion criteria. The review considered studies which involved the ingestion of one or more doses of homeopathic substances (ultra molecular dilutions).

Studies were considered which used substances prepared according to national homeopathic pharmacopoeias or other explicit protocols e.g. the British, French, German, or US Homeopathic Pharmacopoeias.

Studies were considered in which substances were administered in ultra molecular dilutions as are used in homeopathic practice.

Exclusion criteria

Studies involving mother tinctures (crude extracts) of substances were excluded. Whilst these are used within homeopathic practice they are more properly

considered as herbal medicines and cannot help to answer any questions about the effects of ultra molecular dilutions.

Studies which involved the administration of homeopathic substances by means other than ingestion were excluded. Some studies have used olfaction as a method of administering a substance and a number of HPTs have been published using other methods.

2.2.3 *Comparators*

Inclusion criteria

Studies which reported comparing homeopathic medicines to identical placebo medicines.

Exclusion criteria

Studies which did not report using identical placebo medicines as a comparator.

2.2.4 *Outcomes*

The primary aim was to determine whether homeopathic medicines produce effects different from placebo. Outcome measures which capture the kind of symptoms which homeopathic theory predicts would occur were sought. These included symptoms in the kinds of categories expected in HPTs and pre-determined symptoms which might be expected from the particular medicine under trial. Guidance on the conduct of HPTs has suggested that symptoms should be recorded in certain key categories, including new, old, changed, cured, existing, intense and striking or uncommon symptoms (Wieland 1997, Riley 1997, ECH 1995). In relation to pre determined symptoms, some authors define a set of true symptoms which are those that experts deem most likely to occur as pathogenetic effects from the test medicine. (based on either previous HPT's or extensive clinical use). In contrast false symptoms are those which are deemed unlikely to be due to the test medicine and therefore are most likely to be due to placebo or random effects.

The number of typical symptoms recorded per participant or in the whole group are then used as an outcome measure. Also, measures which offered some kind of global assessment of whether a participant displayed a proving reaction or not were sought.

Given the importance and remaining uncertainty over the nature and magnitude of placebo effects as discussed in chapter one, a secondary aim was to determine the typical size of placebo effects in HPT's.

Inclusion Criteria

The review considered studies which reported any outcome relating to symptoms experienced at any point in the duration of the trials. Methods for recording symptoms include the daily completion of an unstructured diary by participants, structured diaries and specially developed proving questionnaires.

Exclusion Criteria

Studies in which outcome were not reported by intervention group.

2.2.5 *Study Designs*

Inclusion Criteria

The review considered randomised controlled trials (RCT's). Such trials are considered to be the best available design for considering the effects of interventions because, if they are properly carried out they best reduce the risk of bias (CRD 2009). This is so because, as Torgerson notes, the RCT provides a simple and elegant design which ensures that, so long as the groups assembled through randomisation are of an adequate size, then there can be confidence that any differences observed between the groups will be a consequence of the intervention, rather than as a result of any other variable, whether known or unknown. (Torgerson and Torgerson 2008). The challenge of constructing meaningful HPTs using RCT design was discussed earlier but in order to find the best and least biased available evidence the review was limited to trials of this design.

Exclusion criteria

-Studies in which there was not a comparator group

-Studies in which the effects of a homeopathic substance were not compared to either placebo or a different homeopathic substance.

- Studies in which participants were not described as randomly allocated to an intervention.

- Studies in which participants were not blinded to treatment allocation

- Studies published in languages other than English for which a full translation was not available.

2.3 Search Strategy

Advice on a search strategy was sought from supervisors, who have expertise in conducting published systematic reviews. Librarians at the University of York and the University of Lincoln were also consulted for advice.

The document published by the centre for reviews and dissemination – CRD guidance for undertaking reviews in health care (2009), was also used, to ensure that current best practice in developing and operating search strategies was followed.

The authors of the published systematic review of the quality of HPT's (Dantas et al 2006) were contacted. These authors considered all papers published between 1945 and 1995. After personal contact they provided details of all of the papers considered for their review. Attempts to contact them again for full details of their search strategy proved unsuccessful. The review by Dantas et al had the advantage of being a multi centre, multi language review, involving twelve reviewers. Although they do not describe the full search strategy in detail, they describe manually searching books, journals and conference proceedings, and using the relevant bibliographic databases (general databases such as MEDLINE and specialist ones such as HomInform).

Given that this review was conducted by a group of international experts in homeopathy, it is likely to be comprehensive and it is very unlikely that, with my limited resources I would find anything new were I to repeat their search.

I therefore decided to use the list of papers supplied by these authors as a starting point and to develop a search strategy only for more recent publications.

Search terms. The advice of specialist searchers and specialist subject librarians is that the following search terms will pick up all papers relevant to homeopathy in general.

1. homeopathy
2. Homoeopathy

   And the following, using relevant truncation symbols

3. Homeop
4. Homoeop
5. Homoop
6. Omeop

Within the homeopathic literature there is a debate about whether the word, which arises from Greek terms homoios and pathos, should have the additional o before the e, but both variants are used widely within the literature. To assist retrieval of non English language papers the variants homoop and omeop were added.

In relation to Provings or homeopathic Pathogenetic Trials, the following search terms cover the possible variants of terminology which are used in relation to this subject

1. Proving (with truncation symbol)
2. prufung
3. pathogenetic trials

B Databases. A range of relevant databases contain information which is relevant to homeopathy.

Different databases have different systems for indexing papers and use a variety of thesaurus and indexing systems to find papers using keywords. Each requires a particular search strategy. The full strategy used to search for the relevant search terms in each of the relevant databases, is contained in Appendices 1 -6.

- MEDLINE, produced by the United States National Library of Medicine, is considered to be the primary source of global information from the international literature, on medicine, healthcare and allied disciplines including complementary therapies such as homeopathy (National Library of Medicine 2008).

- AMED is a specialist database with comprehensive coverage of Complementary medicine.

- EMBASE covers a wider range of journals than MEDLINE, the overlap with MEDLINE is only some 40%. EMBASE also has the merit of being a specifically European database, an important consideration given that much homeopathic literature originates in Britain, Germany, France and Italy.

- CINAHL is the only one database which has a specific index term relating to HPTs – the MESH heading homeopathic Provings is used.

- Hominform is a specialist homeopathic database constructed in 1988 by the Academic Departments of Homoeopathic Medicine of Glasgow Homoeopathic Hospital.

- LILACS (Latin American and Caribbean Health Sciences Literature). This database includes papers published in Spanish and Portugese. All titles, descriptors used for indexing, and some abstracts, are also translated into English. It is known that limiting searches to English language papers can introduce language bias. (Egger et al 1997). Whilst the large databases such as MEDLINE have some non English language journals, the numbers are small. Moher et al (2003) consider the inclusion of papers in languages other

than English to be especially important in systematic reviews of CAM's. Certainly homeopathy is used more widely in some parts of the world, than in Europe and North America. This includes Latin America and ideally it would be important to include LILACS in any search strategy relating to this subject. Searching for trials in LILACS has been demonstrated to improve systematic reviews by finding papers which meet inclusion criteria, but are not found using other database searches in significant numbers of published reviews (Manriquez 2008).

Given these factors it is very likely that a full search including LILACS would reduce bias and would result in the retrieval of additional studies meeting the inclusion criteria. However, the practicalities of an unfunded project meant that a full LILACS search was not feasible. Instead a search of LILACS using English language terms only was undertaken.

The following databases were therefore searched, where possible for the period from 1996 to date.

Table 2.1 Databases used in the search, with date ranges

Database	Date Range
OVID MEDLINE	January 1996- May 2009 (week 4)
AMED	January 1996- May 2009 (week 4)
EMBASE	January 1996- May 2009 (week 4)
CINAHL	January 1996- May 2009 (week 4)
HOMINFORM	All dates
LILACS	January 1996- May 2009 (week 4)

2.4 Stages of the Search

The search was carried out in three stages

Stage one: The full search was carried out as described above. The results of all but one of the database searches were imported into a Refworks database. The full set of references from the aggregated searches was then de-duplicated. It was not possible to import results from The Hom Inform database into Refworks other than via the transfer of each individual data entry; nor was it possible to limit the HomInform search to the relevant date range The HomInform results were therefore screened separately.

The results of hand searches and personal contact with authors were added to the full list.

Stage Two. Screening. Where possible abstracts were screened to determine whether papers met the inclusion criteria. Where a clear judgement could me made papers that did not meet the criteria were rejected. Any papers which met the criteria, or over which there was uncertainty, or for which abstracts were unavailable were obtained in full.

Stage Three. The full papers were then screened, and judgements made about which papers met the inclusion criteria. A second author was asked to verify the screening process. This did not take place before the submission deadline but will be completed shortly after, to assist a process of submission for publication in a journal.

Table 2.2 Algorithm for screening of studies for eligibility

	Assessment	Comment
Randomisation	Yes Unclear No ⇩ EXCLUDE	
Age – 18+	Yes Unclear No ⇩ EXCLUDE	
Interventions Name and strength (potency) of homeopathic medicine (exclude mother tinctures) Placebo described Frequency and numbers of doses described	Yes Unclear No ⇩ EXCLUDE	
Outcomes: Clear quantitative outcomes for both intervention and comparator groups	Yes Unclear No ⇩ EXCLUDE	
Language: Full text of report of trial available in English	Yes Unclear No ⇩ EXCLUDE	

2.5 Data Extraction

All data was extracted by the primary author (JR). The methods used for each trial were noted, including the method of randomisation. Methods of recruitment and details of participants were noted, including details of any methods used to determine the state of health of participants at the point of recruitment. The nature of the intervention was described, including full details of the name and potency of the homeopathic medicine, along with the source and details of the method of preparation of the medicine. Details regarding the nature and preparation of placebo were also described. Full details of primary and secondary outcome measures were described.

CRD note that the extraction of data is linked to assessment of study quality in that both processes are often undertaken at the same time. (CRD 2009).An example of the form that was developed for this review is provided in table 2.1 and it will be seen that data and quality were assessed at the same time on a two part form with separate areas for each.

Primary trial author and year of publication

Methods

Participants

Interventions

Outcomes.

Notes

Risk of bias

Item	Authors Judgement	Description
Adequate sequence	YES/NO/UNCLEAR *	
Allocation concealment	YES/NO/UNCLEAR*	
Blinding	YES/ NO/UNCLEAR*	
Incomplete outcome data	YES/NO/UNCLEAR*	
Screening of volunteers for sensitivity	YES/NO/UNCLEAR*	

*For full details of the criteria used to make these judgements see appendix 2.1

Table 2.1 Example Data extraction and risk of bias form.

2.6 Strategy for assessing study validity.

The terms methodological quality and study validity are often used interchangeably. Quality may refer to both internal and external validity. Internal validity refers to the degree to which results are likely to approximate to the truth (CRD 2009). For this, the design and conduct of the study as well as the analyses of results must be as free from bias as possible. External validity - the degree, to which the results of a study can be applied in other settings, relates to the relevance of the populations, interventions and outcome measures

One of the most significant ways in which findings can be more reliably assumed to be closer to the truth is through the reduction of bias. With studies involving human subjects there are a range of types of bias which can skew the results of research, often significantly.

The sources of such bias can be summarised as arising from subjects, researchers and measurements (Peat et al 2002) and the types of bias have also been categorised in the following way (Khan et al 2003)

1. Selection (or allocation) bias. This refers to systematic differences between comparison groups in some measure which may affect the outcome being measured.
2. Performance bias. This refers to differences in the care or interactions between clinicians and/or researchers and the different comparator groups.
3. Measurement bias (also known as detection bias). This refers to differences in the measurement of outcomes between the comparator groups
4. Attrition bias. (also known as exclusion bias). The bias can arise if features of the intervention (such as side effects) lead to a higher drop out/ withdrawal rate from one of the comparator groups.

Individual studies are subject to all of these types of bias. It is important, in trying to make realistic assessments of the efficacy and effectiveness of health care interventions to make some kind of assessment of quality in relation to these aspects

since several studies have indicated that health care trials of poor methodological quality offer exaggerated estimates of treatment effects (Schultz et al 1995)

One of the key aims of systematic reviews is also to reduce bias.(Khan et al 2003) and, as well as quality tools for the reporting of individual trials there are a range of tools available to assessors to assist in the assessment of trials by reviewers.

There are a number of reasons for assessing the level of bias and the quality within the primary studies which are conducted in any review. The inclusion of the results of a series of poor quality and biased studies in any meta- analysis or quantitative analysis which follows a review may of course simply amplify the bias. Other reasons suggested for assessing quality of primary studies include:

- Determining a minimum threshold of quality in terms of study design which will form the basis of selection for a review
- To explore quality differences as an explanatory framework in relation to the heterogeneity of study results
- Weighting study results according to quality in meta- analyses

(Khan et al 2003)

Many methodological quality instruments and scales have been developed to assess these parameters in trials. A widely used scale is the one developed by Jadad and colleagues at the University of Oxford. This offers a simple scoring system with five possible points awarded for elements which appear in the published version of the trial. Points are scored if a study is described as randomised, described as double blind and if it offers a description of withdrawals and dropouts. Its simplicity and relevance for the task has led to it becoming the most widely used instrument, and as of 2008 the original paper had been cited over 3000 times (Olive et al 2008).

It is often used not only as an assessment of some of the key elements of validity but to provide a cut off point for using trials in reviews and meta- analysis. Systematic reviewers will often use a score of 3 as a minimum for including trials in reviews.

(Berger 2006). The JADAD scale has the merits of simplicity, credibility and a clear focus on those aspects of trial design which are most likely to reduce bias. The elements identified in the JADAD scale are important but there are criticisms of scales and scoring systems in relation to validity. The guidance from the Centre for Reviews and Disseminations suggests that "*The use of scales with summary scores to distinguish high and low quality studies is questionable and not recommended*" (CRD 2009). They suggest instead that individual aspects of methodological quality should be considered when assessing quality and at the stage of data synthesis. The impact of quality on findings can then be assessed by sub group analysis.

Whilst it is important to consider all of the aspects of quality which consensus methods have agreed to be important for clinical trials, it is also important to be sensitive to the specific elements of quality which might be unique to the types of studies under consideration. It was noted in the background chapter that HPT's pose a unique set of challenges and questions. Researchers have noted that there has been a lack of consensus and a wide variety of methodologies employed in the design and conduct of HPT's (Wieland 1998)

Expert groups, and those who have conducted numbers of HPTs have made several sets of recommendations designed to improve the quality of HPT's. (Wieland 1997, ECH 1995, Riley 1997) In their systematic review of the quality of HPT's Dantas et al, building on earlier work, developed a methodological quality index which included the usual features of randomisation and blinding but also included a score relating to the " criteria for the selection of effects ".(Datas et al 2007)

As noted in the background chapter the issue of susceptibility is crucial in the design of HPT's and methods for dealing with this issue should be reported in HPTs. According to homeopathic theory this would seem to be the factor which would have the greatest impact on outcomes.

Therefore, in this review, the quality of each included study was assessed in relation to randomisation and allocation concealment, blinding, strategies for dealing with missing outcome data, and strategies for determining the sensitivity of participants to the medicine used in the trial. A risk of bias form was completed for each study to

record details and a judgment of yes, no, or unclear was made in relation to the achievement of each of these elements. (Table 2.1)

Full details of the criteria which were used to make these judgements are contained in appendix 7

2.7 Strategy for data synthesis

Synthesis involves the drawing together of disparate elements, and the summarising of the findings of all of the studies included in the review. In evidential terms it is viewed as preferential if some kind of quantitative synthesis can be arrived at using statistical techniques such as meta- analysis. This is only possible where the individual studies do not have methodologies and /or results which are too disparate or heterogeneous. If the studies are too diverse to allow for meta-analysis then a narrative summary only should be provided.

For this review a narrative synthesis was developed initially. Tables and graphs comparing studies in relation to: patients, interventions, comparators, and outcomes were constructed and follow in the results section. A discussion and analysis of the design and findings of the included studies follows.

2.8 Heterogeneity.

It is important to consider heterogeneity when comparing studies. Heterogeneity refers to the amount of difference or diversity between different studies in terms of reported outcomes. It can arise for various reasons - it may be related to differences in study populations, in the methodology of the studies and in the outcome measures used, as well as to chance variation. (Higgins et al 2003). The clinical heterogeneity which may arise for these reasons can be assessed and quantified statistically. The I^2 statistic, for example, calculates the amount of variability in the estimates of effect that can be attributed to heterogeneity other than chance. Guidance has been published which suggests that values up to 40% may be unimportant, up to 60% deemed as moderate and over 75% considerable. (Higgins and Green 2008)

One of the advantages of graphical displays such as forest plots is that they offer clear visual evidence of the level of heterogeneity – where there is little overlap between confidence intervals this is a sign of high heterogeneity. Forest plots were constructed for this review as part of a meta-analysis.

2.9 Meta analysis

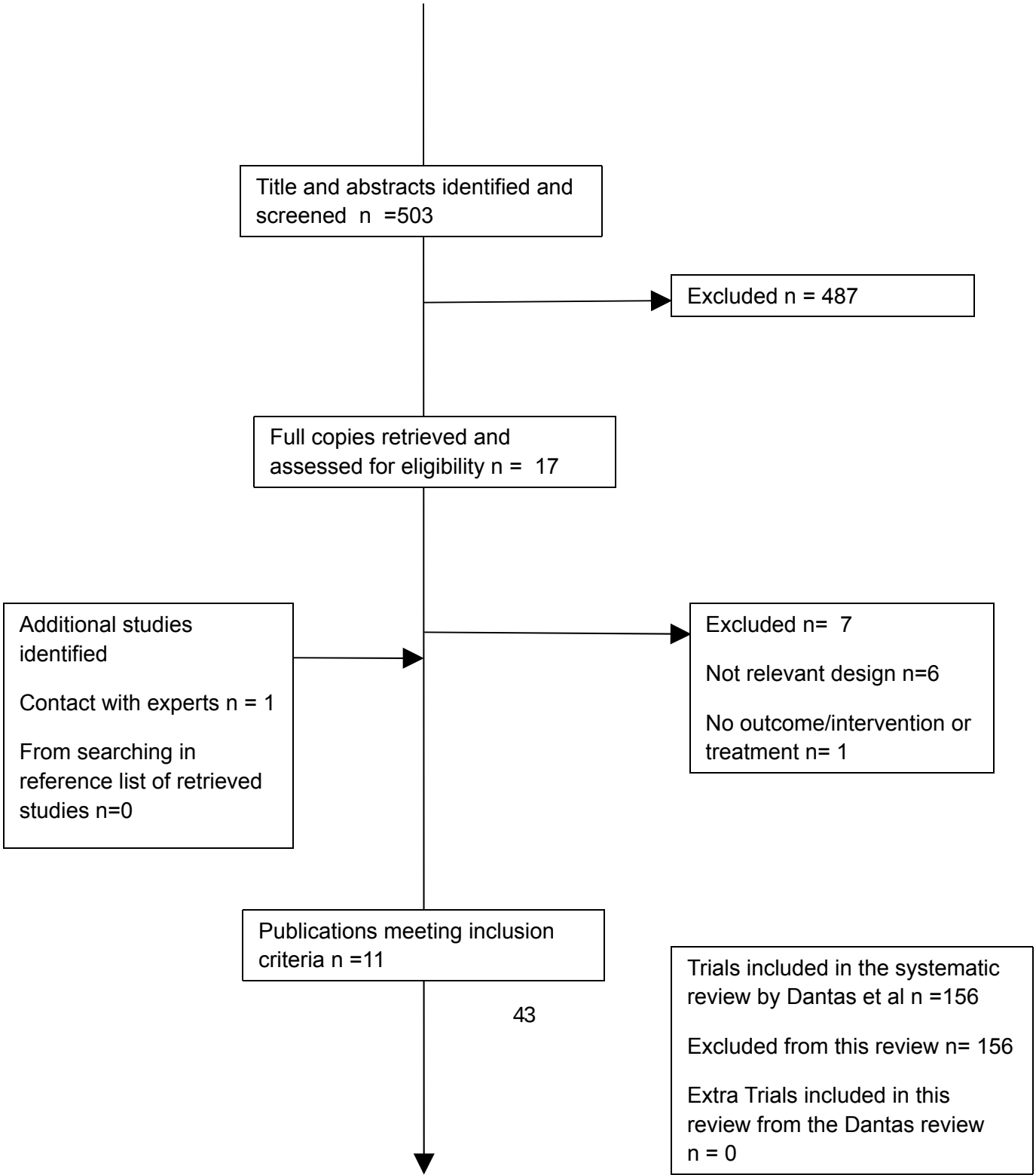
The amount of heterogeneity also determines the methods used to treat data in any meta-analysis. If heterogeneity does not exist or is limited a statistical meta-analysis can be performed using a fixed effects models (which assume that all studies under analysis are estimates of a single or very similar population). Where there is clear heterogeneity, meta- analysis can still be performed but using random-effects models to account for the greater uncertainty in the individual estimates In this review it was possible to pool some data for two of the key outcome measures: overall proving reactions and symptoms typical of the test medicine. Forest plots were constructed, using both fixed and random effects.

Where data is pooled from trials measuring the same outcomes on the same scale straightforward mean differences can be calculated; these are often weighted according to trial size or levels of statistical variance between trials, to arrive at weighted mean differences (WMDs). In this review effect sizes in the form of the alternative standardised mean differences (SMDs), and 95% confidence intervals, have been calculated where possible. SMDs allow for comparisons across different outcomes measures relating to the same variables of interest, (CRD 2009) and this is appropriate because the trials assessed for this review used a variety of different measures.

CHAPTER THREE RESULTS

Database	Date Searched	Number of papers retrieved	Full Details
OVID Medline	31/5/09	44	Appendix 1
AMED	1/6/09	323	Appendix 2
Cochrane register of controlled trials	14/6/09	143	Appendix 3
EMBASE	17/6/09	38	Appendix 4
CINAHL	17/609	56	Appendix 5
Hominform	12/6/09	503	Appendix 6

After de-duplication a total of 503 papers were left for screening.Screening of the abstracts and/or full texts of these papers against the inclusion criteria for this review led to the exclusion of 487 of them. 17 papers were retrieved and read in full. One of these was excluded (Koster et al 1998) because it was a report mid way through a trial, with no results. Efforts to contact the author and searches for any published version of the final outcomes were unsuccessful. Six other papers were excluded because they did not use appropriate design with random allocation of participants to intervention or placebo groups. Personal contact with the author of two of the papers which had met the inclusion criteria, led to the inclusion of a further paper recently published in German – the author provided a full translation. Four of the papers took the approach of combining reports from two separate trials into a single paper, meaning that the 11 papers included in the review covered a total of 15 trials. All of the papers were published in journals.


Title and abstracts identified and screened n =503
Excluded n = 487
Full copies retrieved and assessed for eligibility n = 17
Additional studies identified
Contact with experts n = 1
From searching in reference list of retrieved studies n=0
Excluded n= 7
Not relevant design n=6
No outcome/intervention or treatment n= 1
Publications meeting inclusion criteria n =11
Trials included in the systematic review by Dantas et al n =156
Excluded from this review n= 156
Extra Trials included in this review from the Dantas review n = 0

Publications included in the reviews

n =11

Number of studies included in the review n = 15

Figure 3.1 Flowchart of included studies.

Summary of included studies.

3.1 Authors There were 8 different primary authors, with multiple studies by Vickers (primary author on two studies) and Walach (primary author of 3 studies and contributing author on 2 others)

3.2 Settings

The studies took place in a variety of countries: 7 in the UK, 3 in Germany, 3 in Italy and 1 in both the UK and Israel.

3.3 Participants

It is clear from table 3.2 (below) that studies involved more females that males. From the available data, 395 participants were female and 182 male, proportions of 68.5% and 31.5%. Walach (2004) did not report on gender and Walach (2001) reported proportions of 49% and 66% female in two different samples, but did not give actual numbers. From the remaining 13 studies, the percentage of females in the sample was over 75% in four of the studies. (Vickers Van Haselen 2001, Brien 2003, Walach 2008, Mollinger 2009). 9 of the 13 studies which reported had more female participants that male. Just one of the studies (Vickers, Mccarney 2001) had over 60% males.

Table 3.2 Gender of participants in included studies

Author	**Female**	**Male**
Wallach 2001[HPT1]	2 samples n unclear Sample A homeopathic community =49% Sample B naive subjects =66%	Sample A =51% Sample B = 34%
Vickers Van Haselen 200 [HPT2]	80 (78%)	24 (22%)
Dominici et al 2006 [HPT 3A] HPT3B	9 (64%) 8 (62%)	5 (36%) 5 (38%)
Fisher 2001 [HPT4A] HPT4B	14 (70%) 11 (55%)	6 (30%) 9 (45%)
Brien 2003 [HPT5]	164 (79%)	42 (21%)
Goodyear 1998 [HPT6]	23 (48%)	24 (52%)
Walach 2004 [HPT7]	Not reported	Not reported
Signorini 2005 [HPT 8A] HPT8B	9 (50%) 11(44%)	9 (50%) 14 (56%)

Walach 2008 [HPT9A] [HPT9B]	17 (100%) 20 (55%)	0 (0%) 16 (45%)
Vickers McCarney 2001 [HPT10]	19 (38%)	31 (62%)
Mollinger 2009 [HPT11]	19 (76%)	6 (24%)
Total	395 (68%)	182 (32%)

Despite it being standard practice to include only healthy adults in pathogenetic trials only one study author reported attempts to define and operationalise healthy participants. Fisher (2001) included only those who scored within one standard deviation of the mean value for UK adults on at least seven of eight subscales of the SF36 short form medical outcomes questionnaire. In addition participants in this trial had to undergo screening for full blood count, plasma urea and electrolytes and return these within the normal range. Brien (2003) used an unspecified questionnaire to screen for stable good health. Walach (2001) used a recently validated personality questionnaire – the Freiburg Personality Inventory (FPI) as part of the trial, but not as part of the inclusion/exclusion process.

All of the trials used inclusion and exclusion criteria in order to try to recruit only adults in stable good health. (see table 3.1 below). It will be seen that whilst each trial used a unique set of criteria there were some commonalities. 5 of the studies excluded participants who were pregnant or nursing and one of these (Walach 2001) extended this to include planned pregnancies as well as current ones. The use of different kinds of medicines was approached in a range of ways. 2 of the trials (Walach 2001 and Fisher 2001) excluded those using recreational drugs though Fisher made an exception for tobacco in regular smokers, and moderate use of alcohol. 9 of the studies adopted explicit exclusion criteria relating to the use of

prescribed medication, though there were variations in the way in which these were worded and again some exceptions were made. Walach (2001), Mollinger (2009) and Dominici (2006) mentioned any use of medication; Vickers and Van Haselen (2001) mentioned regular use; Fisher (2001), Brien (2003) and Goodyear (1998) referred to the use of medication in the previous 4 weeks; for Signorini (2006) it was 2 months, and for Vickers and McCarney it was only concurrent use during the trial that was excluded. In relation to the specific use of Homeopathic medicines other than the trial medicine this was mentioned in 11 of the 15 trials but there were again differences in the period exclusion. Vickers and VanHaselen (2001) and Walach (2008) simply referred to regular use. Dominici (2006) excluded anyone using homeopathic medicines in the last 6 months, whilst for Brien (2003) and Goodyear (1998) it was just 4 weeks. For Walach (2004) it was 10 weeks.

Table 3.3 Comparison of exclusion criteria used in individual trials in relation to health status of participants.

Trial	Exclusion Criteria
Walach 2001 HPT1	▪ Use of recreational drugs or any conventional medication ▪ Abuse of alcohol, or pharmaceutical drugs ▪ Pregnancy, (present or planned), or nursing ▪ Stress, Irregular life
Vickers 2001 HPT2	▪ Self report of any perceptible illness ▪ Regular use of any form of homeopathic, herbal or medical drugs (stable use of contraceptive pill for at least one year allowed) ▪ Formal training in or explicit knowledge of homeopathy ▪ Anticipated major life change during the trial ▪ Pregnancy or nursing ▪ Participation in other HPTs or clinical trials in previous 3 months ▪ Lack of mental capacity to understand the nature or consequences of the trial
Dominici 2006 HPT 3A	▪ Use of contraceptive pill or any drugs ▪ Elective medical treatment during the trial period ▪ No homeopathic remedy within the previous 6 months.

	▪ Any mental pathology or chronic physical pathology in the judgement of the participant or the trial supervisor.
Dominici 2006 HPT3B	▪ As for HPT3A (above)
Fisher 2001 HPT4A	▪ Current medical treatment ▪ Prescription only medication in previous month ▪ Use of recreational drugs (except alcohol in moderation and tobacco if established smoker. ▪ Surgery in preceding 6 weeks ▪ Pregnancy, breast feeding. ▪ Values more than one standard deviation outside the mean value on the SF36
Fisher 2001 HPT4B	▪ As for HPT4A (above)
Brien 2003 HPT5	▪ Any medication use (conventional, herbal or homeopathic) in the previous 4 weeks (contraceptive pill allowed) ▪ Illness during trial requiring excluded medication ▪ Current pregnancy/nursing.
Goodyear 1998 HPT6	▪ Any medication use (conventional, herbal or homeopathic) in the previous 4 weeks ▪ Self report of any perceptible illness
Walach 2004 HPT7	▪ Use of any homeopathic medicine or participation in HPT in previous 10 weeks. ▪ Any organic pathology
Signorini 2005 HPT8A	▪ Chronic disease in the previous year. ▪ Hospital treatment(past 6 months), vaccination (past 3 months), dentistry (past month) ▪ Medication in last 2 months (6 months if homeopathic meds in potencies above 200C
Signorini 2005 HPT8B	▪ As in HPT8A (above)
Walach	▪ Use of homeopathic medicine in previous 4 – 12 weeks (depending on potency)

2008 [HPT9A]	▪ Regular use of conventional, recreational or homeopathic medicines. ▪ Any acute or chronic medical condition
Walach 2008 [HPT9B]	As for HPT9A (above)
Vickers 2001 [HPT10]	▪ Self report of any perceptible illness ▪ Concurrent use of any form of homeopathic, herbal or medical drugs (stable use of medication and stable illness was allowed following an assessment by the study doctor) ▪ Anticipated major life change during the trial
Mollinger 2009 [HPT11]	▪ acute illness, or chronic disease necessitating regular medication ▪ intake of conventional medication except contraceptives, ▪ pregnancy or nursing, ▪ extraordinary strain from family or job demands, ▪ intake of any homeopathic remedies at time of trial, ▪ intake of homeopathic remedies at C30 up to 4 weeks previously, intake of homeopathic remedies at C200 up to 2 months previously, intake of homeopathic remedies at C1,000 or higher up to 3 months previously

Other authors made distinctions, dependent on the strength of the homeopathic medicine. So Signorini excluded us of any homeopathic medicine in the previous 2 months, but this was extended to 6 months if the strength of the homeopathic medicine was 200c or higher. In a similar way Walach (2008) and Mollinger (2009) had varying exclusion periods, ranging from 4 weeks up to 12 weeks for the higher strength homeopathic medicines. As a corollary to the exclusion of those using homeopathic medicines for treatment, 2 authors excluded those who had participated in other HPT's in the past 3 months or 10 weeks (Vickers and Van Haselen 2001, Walach 2004 respectively)

Vaccination and dentistry were explicitly excluded by just one author (Signorini 2006). Vickers and Van Haselen (2001) were the only authors to explicitly refer to mental capacity

3.4 Susceptibility to a homeopathic medicine.

None of the included studies report any methods for screening participants for sensitivity or susceptibility to the specific medicine used in the trial.

3.5 Interventions and comparators

Of the 15 studies, 13 studies compared single homeopathic medicines with identical placebo medicines (an identical sugar pill or liquid solution). 2 studies (Walach 2008 and Mollinger 2009) involved three arm trials which compared 2 separate homeopathic medicines with placebo. Table 3.4 lists the schedules for individual studies.

9 of the studies used a fixed dosage schedule with participants taking medicines twice daily (Walach 2001; Brien 2003; Goodyear 1998) or 3 times daily (Vickers, Van Haselen 2001; Fisher 2001, Vickers, McCarney 2001) for the duration of the trial. 5 of the papers, covering 6 separate trials report flexible schedules with participants instructed to stop use of medication if symptoms occurred.

10 of the trials involved the use of homeopathic medicine in a 30C potency, and 4 used medicines in a 12C potency. No other potencies were used. Belladonna was chosen as the test medication in 3 trials (Brien 2003, Goodyear 1998 and Walach 2001). The remaining trials all used different, single medicines.

Table 3.4 Interventions. Medicines and dosages.

Walach 2001 HPT1	Belladonna 30C	Twice daily
Vickers 2001 HPT2	Mercurius 12C	Three times daily
Dominici 2006 HPT 3A	Etna Lava 30C	Three times daily but stop if new symptoms appear
Dominici 2006 HPT3B	Hydrogenium Peroxidatum 30C	Three times daily but stop if new symptoms appear
Fisher 2001 HPT4A	Acidum Malicum 12C	Three times daily
Fisher 2001 HPT4B	Acidum Ascorbicum 12C	Three times daily
Brien 2003 HPT5	Belladonna 30C	Twice daily

Goodyear 1998 [HPT6]	Belladonna 30C	Twice daily
Walach 2004 [HPT7]	Cantharis 30C	Twice daily
Signorini 2005 [HPT8A] Signorini 2005 [HPT8B]	Piper Methysitcum 30C/ Plumbum Metallicum 30C	Four times daily, increased to six after third day but stop if strong symptoms
Walach 2008 [HPT9A] Walach 2008 [HPT9B]	Ozone 30C Ozone 30C	Several times daily at discretion of Participant, for no more than 3 days
Vickers 2001 [HPT10]	Bryonia 12C	Three times daily
Mollinger 2009 [HPT11]	Natrum Muriaticum 30C/ Arsenicum Album 30C	Single dose day 1. Twice daily day 2 onwards or until symptoms appear.

Table 3.5 Outcome measures used in included trials

Trial	**Outcome measure**	**Timing**	**Dimensions measured for primary and secondary outcomes**
Walach 2001 [HPT1]	Unstructured diary	Daily	Number, frequency and intensity of sx
Vickers 2001 [HPT2]	Questionnaire with 5 sx deemed true and 5 deemed false for the medicine	Daily	Frequency of each sx Intensity on 3 point scale
Dominici2006 [HPT 3A]	Structured diary	Daily	Existing symptoms increasing in intensity/duration Previous sx that had not occurred for at least 1 yr. Current sx that disappeared during HPT New sx unfamiliar to participant Exceptional sx
Dominici 2006 [HPT3B]	Structured diary	Daily	As HPT3A above
Fisher 2001[HPT4A]	Structured and unstructured	Daily	Severity, frequency, duration, modality

	dairies		and causal relationship to treatment.
Fisher 2001[HPT4B]	Structured and unstructured dairies	Daily	Severity, frequency, duration, modality and causal relationship to treatment.
Brien 2003 [HPT5]	Questionnaire with 5 sx deemed true and 5 deemed false for the medicine	Daily	Proving reactions: defined as 2 true sx on at least 2 consecutive days with no more than one false sx in 21 day study period.
Goodyear 1998 [HPT6]	Questionnaire with 5 sx deemed true and 5 deemed false for the medicine	Daily	Proving reactions: defined as 2 true sx on at least 2 consecutive days with no more than one false sx in 14 day study period.
Walach 2004 [HPT7]	Unstructured diary	Daily	Symptoms deemed typical of the medicine.
Signorini 2005 [HPT8A]	Unstructured diary	Daily	Selected notes containing sentences describing symptoms which were very unusual or never happened before.
Signorini 2005 [HPT8B]	Unstructured diary	Daily	As HPT8A above
Walach 2008 [HPT9A]	Unstructured diary and interview by supervisor	Daily	Particular attention paid to symptoms never experienced before.
Walach 2008 [HPT9B]	Unstructured diary and interview by supervisor	Daily	Particular attention paid to symptoms never experienced before.
Vickers 2001 [HPT10]	Questionnaire	Any point during or after 2 x 1	Participant guess as to which of two medicines taken was homeopathic medicine and which placebo.

		week trial periods	Confidence in judgement on 3 point scale
Mollinger 2009 HPT11	Unstructured diary and interview by supervisor	Daily	Judgement by homeopathy expert as to whether symptoms were typical or not for the homeopathic medicine taken.

Outcomes

As illustrated by Table 3.5, a range of instruments were used to capture the experience of symptoms by participants during the pathogenetic trials. 12 of the 15 trials involved the completion by participants of structured and/or unstructured diaries on a daily basis, allowing for the recording of any symptoms which occurred. 4 papers reported trials in which certain symptoms were pre-specified as typical of the medicine being tested; participants were then asked to record whether a range of symptoms typical and untypical for the medicine were experienced. Outcomes in these trials involved comparing numbers of typical and untypical symptoms across the verum and placebo groups (Goodyear 1998, Vickers, Van Haselen 2001, Brien 2003, Walach 2008). Fisher (2001) developed a pathogenetic index which was adapted from an index for judging the causality of adverse drug reactions. (Naranjo 1981). The index offered a score based on judgements of both participants and supervisors. Only 1 trial (Signorini 2006) was designed with outcome measures to capture symptoms by the general categories used in homeopathic literature (e.g. new, old, unusual, cured, intense,)

3.7 Design

All of the trials included in this review were randomised trials. Authors reported using randomisation to allocate participants to homeopathic medicine or identical placebo medicine. Authors are divided with regard to the use of parallel or crossover designs for these types of trials. Walach (2001), Fisher (2001) and Vickers, McCarney (2001) report crossover designs. The remaining trials were all designed as parallel group

trials. Table 3.2 shows the range of baseline, intervention and follow up periods, of the trials.

Table 3.6 Timescales of included trials.

Trial	Run in period	Intervention period	Follow up period
Walach 2001[HPT1]	2 wks observation	2 x 1 wk (crossover) with 1 wk run in and 1 wk washout	-
Vickers, Van Haselen 2001 [HPT2]	1 wk placebo	I wk with 1 wk placebo run out	-
Dominici 2006 [HPT3]	2 wks observation	2 days or until symptoms appeared	2 months
Fisher 2001 HPT4A AND HPT4B	Unspecified period – observation only	4 x 1 wk (crossover)	1 wk – extended to 3 if symptoms continuing (report of any further symptoms after next 60 days)
Brien 2003 [HPT5]	1 wk placebo	2 wks	1 wk
Goodyear 1998 [HPT6]	None	2 wks	-
Walach 2004 [HPT7A AND HPT7B]	1 wk observation	2wks	-
Signorini 2005 [HPT8A AND	1 wk observation	1wk	1wk

HPT8B			
Walach 2008 HPT9A AND HPT9B	1 wk observation	3 days or until symptoms appeared	2 wks
Vickers,McCarney 2001 HPT10	None	2 x 1 wk (crossover)	
Mollinger 2009 HPT11	None	2 days	2 days

3 of the 15 trials had no run in period, and did not measure any outcomes during a baseline period. Most of the trials, (10 from 15) had a run in period. 1 week was the most common length: 3 authors used a 1 week observation period in 5 separate trials; and 2 authors used a 1 week run in period with the use of placebo. 2 authors used a longer 2 week run in period for observation. One of the trials discussed an observation period between recruitment and the start of the trial but did not specify the length of this.

Turning to the length of intervention periods in the included trials, the majority (11 of the 15) used 1 or 2 weeks. 4 trials used a 2 week intervention period, a single trial used a 1 week period, with a further trial using a one week active intervention period followed by a 1 week placebo run out. The 5 crossover trials all used one week intervention periods interspersed with 1 week washout periods.

3.8 Risk of bias in included studies

Methodological quality was assessed using the key criteria suggested by the Cochrane Collaboration Tool for Assessing Risk of Bias (Higgins JPT, Green S, 2008) and as outlined in the methods section. Full details for each study are in the risk of bias tables which form part of the table of included studies. Appendix 7

3.9 Sequence generation and allocation concealment

In 14 of the 15 included trials details of the sequence generation and the concealment of randomisation were reported and explained in sufficient detail to merit a positive judgement in relation to these features. Just one of the trials (Vickers, McCarney 2001) did not provide sufficient detail regarding the method of randomisation to be able to make a clear judgement about the method of sequence generation in relation to allocation to groups.

3.10 Blinding and predicting assignment.

All of the trials involved procedures for double blinding and report techniques for blinding trial supervisors and participants. The issue of verification of blinding was touched on previously. The ability of participants to guess which intervention they have received is sometimes used as a measure of blinding. This makes sense in certain circumstances, particularly at the beginning of trials. However, if participants receive interventions which do have significant effects (beneficial or adverse) which provide signals to them that they are receiving some kind of active intervention then logically they are more likely to guess their assignment, if the comparison is between placebo and verum. It is known that when side effects are prominent patients can be unblinded in clinical trials (Hertzman and Feltner 1997). Walach (2008) notes that

HPTs are designed to actually elicit what would be deemed as side effects in a clinical context, and that using participants judgements about assumed group assignment cannot then sensibly used to verify blinding.

The alternative method of asking people who were not trial participants to guess the difference between placebo and verum preparations was used by Vickers and Van Haselen (2001) and Brien 2003 who used an independent group from the MRC clinical trials unit to verify indistinguishability of placebo and verum preparations.

Fisher (2001) reports two trials in which correct guessing of treatment allocation (placebo vs. verum) by volunteers was 48% and 50% respectively. It is not fully clear at which point in the trial this guessing was carried out.

Vickers and McCarney (2001) actually used participant's guesses about group assignment as the primary outcome measure to determine whether Homeopaths can detect homeopathic medicines. In their study 70 homeopaths were randomised of whom 50 completed the trial. 60% correctly identified the bottle containing the homeopathic remedy Bryonia 12C.

3.11 Follow up and exclusions

The issue of withdrawal and loss to follow up from trials of health interventions is important. If data is excluded from analysis because of loss to follow up, there may be systematic differences between such data across the comparator groups. If differences in interventions and outcomes exist they may lead to attrition bias. A clear example is one in which high numbers of people withdraw from an intervention group due to adverse events. If results are analysed only on the basis of those who complete the trial, such an analysis is likely to favour the intervention group in many trials. A trial of good quality should have clear processes in place for gathering full data as far as possible, and thus minimising attrition bias and its effects (Wright and Sim 2003). Intention to Treat (ITT) analysis is recommended so that data on trial participants is analysed according to the group to which they were randomised. Imputation of missing data can be made, based on explicit criteria. As a minimum trial authors should report numbers and proportions lost to follow up across the

different comparator groups so that reviewers can make judgements about whether the number and pattern of participants lost to follow up us likely to have led to bias

In the trials considered for this review, a total of 692 participants were enrolled in studies. Total loss to follow up was 118 and 574 participants completed the trials, meaning that there was a total 17% loss to follow up. Table 3.7 lists the numbers of and reasons for , loss to follow up in included studies.

It can be seen from this table that the most common reason for loss to follow up is failure to complete the daily diary or questionnaire adequately. Adverse events were explicitly noted as a reason for loss to follow up in 2 trials by Signorini (2006). Adverse events are discussed in more detail in the next section.

Table 3.7 Loss to follow up in included studies.

Trial	No of participants enrolled	Number lost to follow up by group (and % of total)	Reasons for Loss to follow up
Walach 2001[HPT1]	118	31(26%) Crossover trial. Timing of withdrawals not specified	10= non return of diairies,21 diaries contained incomplete data.
Vickers, Van Haselen 2001[HPT2]	118	Verum n=8 (13%) Placebo n=6 (10%)	Authors report that all occurred in placebo run in week and were balanced across two groups which were comparable at baseline
Dominici 2006[HPT3]	16	0 (0%)	N/A
Fisher 2001[HPT4A]	20	n=3 (14%)	Failure to attend interview twice (2)

		Crossover trial. Timing of withdrawals not specified	Failure to complete diary(1)
Fisher 2001[HPT4B]	20	n=5 (25%) Crossover trial. Timing of withdrawals not specified	Request of participant (3) Failure to complete diary(2)
Brien 2003[HPT5]	234	Verum n=25 (20%) Placebo n=22 (17%)	Too busy (placebo =9, verum =10) Lost medication (p = 1 v =4) Non compliance (p =1 v =3) Breeched exclusion criteria (p=1 v=2) Unknown (p = 10, v =6)
Goodyear 1998[HPT6]	60	Verum n=10(33%) Placebo n=3(10%)	Authors unable to determine
Walach 2004[HPT7]	11	unclear	N/A
Signorini 2005[HPT8A]	32	Verum n=3(19%) Placebo n=5 (31%)	Verum -Adverse reaction (1) – pain and swelling of knee. Placebo:Tonsillitis(1) Use of anti-inflammatory drugs (1) Absence of baseline history (1) code breaking(1)
Signorini 2005[HPT8B]	31	Verum n=7 (53%) Placebo n= 5	Verum – adverse event – neuralgia (1) withdrawals, no reason given (6) Placebo (same placebo group used for trial HPT8A and HPT8B):Tonsillitis(1)

		(31%)	Use of anti-inflammatory drugs (1) Absence of baseline history (1) code breaking(1)
Walach 2008[HPT9]	36	0 (0%)	N/A
Walach 2008[HPT9]	17	0 (0%)	N/A
Vickers, McCarney 2001 [HPT10]	70	20 (29%) Crossover trial.	Failure to return questionnaire. Concomitant disease. Change in eligibility. Withdrawal of consent.
Mollinger 2009[HPT11]	25	0(0%)	N/A

3.12 Adverse events. See table below.

Table 3.8 Adverse events in included studies

Trial	Number of adverse events	Comments
Vickers, Van Haselen 2001 [HPT2]	N= 62 (in 48 different participants) During placebo run in n=22 Verum n=22 Placebo n=19	No significant difference between verum and placebo groups.
Fisher 2001 [HPT4a]	No serious adverse events	1 participant withdrawn due to intercurrent illness. (fever, headache runny nose). On breaking code found to be in placebo group.
Fisher 2001 [HPT4B]	No serious adverse events	1 participant withdrawn due to intercurrent illness. (itching, poor concentration, muscular pain). On breaking code found to be in placebo

		group
Brien 2003[HPT5]	37 (in 253 participants)	Two required hospitalisation (1 in verum and 1 in placebo group) No significant difference between verum and placebo groups
Signorini 2005 [HPT8A]	1 (in 16 participants) in verum group 0 (16 participants) in placebo	Pain and swelling of knee.
Vickers, McCarney 2001[HPT10]	3 (in 70 participants)	All in washout period (no intervention) following verum

Adverse events can be a significant reason for loss to follow up in trials. A total of 106 separate adverse events were reported by different participants in five trials considered for this review. The remainder of the trials did not report on this item.

Vickers and Van Haselen (2001) reported that adverse events which led to the use of other medication or to withdrawal from the study were recorded . Causal relationships between such adverse events and the study medication were assessed. They reported 62 adverse events in 48 subjects but no significant differences between placebo and verum and no evidence of adverse events in those taking homeopathic mercury. Fisher (2001) reported no serious adverse events. Brien (2003) defined adverse events as those symptoms which required medication other than the study medication and also true symptoms that led to withdrawal from the study. They reported 37 AEs including two serious issues requiring hospitalisation (one in the verum and one in the placebo group). They reported no significant difference in adverse events between verum and placebo or between those displaying a proving reaction and those not. Signorini et al (2005) reported just one adverse reaction, occurring in the verum group. Vickers, McCarney et al (2001) reported adverse events in three participants who all continued with the trial, two to completion. In each case the event occurred during a washout period following the

period of taking verum. The other 7 papers included this review do report the occurrence or management of adverse events.

None of the German studies discussed the issue of adverse events and only 2 of the larger UK studies noted any significant numbers of adverse events. The aim of pathogenetic trials is to produce pathogenetic symptoms and it might be expected that significant pathogenetic effects might be classed as adverse events. A clear definition of an adverse event and the difference between a pathogenetic symptom and an adverse event is recommended in ECH guidelines and this would help the clarity of reporting.

It is notable that the study which reported the highest number of adverse events (Vickers and Van Haselen 2001) reported that the study nurse specifically asked participants at the end of each week about the occurrence of adverse events.

The two studies which reported significant numbers of adverse events (Vickers and Van Haselen 2001, Brien 2003) both took the approach of recording as an adverse event that which led to the use of medication (other than the study medication) or to withdrawal from the trial. Neither of these studies reported a significant difference between verum and placebo groups in the occurrence of adverse events. This could be taken as evidence that adverse events were unlikely to have biased the outcomes in significant ways. It could also be taken as a form of evidence that homeopathic medicines do not lead to significant pathogenetic effects any more than placebo interventions do.

3.13 Screening for sensitivity to the test medicine.

This item was noted above in relation to the discussion of inclusion and exclusion criteria. It is deemed a critical item which will affect the ability of trial designs to detect outcomes. It is therefore included also as one of the items used to judge the quality of trials.None of the included studies report any methods for screening participants for sensitivity or susceptibility to the specific medicine used in the trial.

3.14 Summary of the overall quality of studied included in the review.

Overall, the studies which met the inclusion criteria for this review were of good quality in relation to methods of randomisation, allocation concealment, and blinding.

(see table 3.9) Methods of dealing with loss to follow up and adverse events were more variable. Clear definitions and methods for checking for adverse events during trials were missing in 8 of the papers included. Five of the trials reported loss to follow up of 20% or greater.

Table 3.9 summary of judgements relating to quality of included studies

Trial	Adequate sequence generation	Allocation concealment	Blinding	Procedures for dealing with incomplete outcome data	Screening of participants for sensitivity to homeopathic medicines
Walach 2001[HPT1]	YES	YES	YES	NO	NO
Vickers, Van Haselen 2001 [HPT2]	YES	YES	YES	YES	NO
Dominici 2006 [HPT3]	YES	YES	YES	NO	NO
Fisher 2001 HPT4A AND HPT4B	YES	YES	YES	UNCLEAR	NO
Brien 2003 [HPT5]	YES	YES	YES	YES	NO
Goodyear 1998 [HPT6]	YES	YES	YES	UNCLEAR	NO
Walach 2004 [HPT7A] AND HPT7B	YES	YES	YES	UNCLEAR	N0
Signorini 2005 [HPT8A] AND HPT8B	YES	YES	YES	NO	NO
Walach 2008 [HPT9A] AND HPT9B	YES	YES	YES	YES	NO
Vickers,McCarney	UNCLEAR	YES	YES	YES	NO

2001[HPT10]					
Mollinger 2009 [HPT11]	YES	YES	YES	YES	NO

3.15 Study results. Primary and secondary outcomes

The primary outcomes and other items of interest ,as discussed in the methods section, are reported here. Authors most commonly reported outcomes using mean values.

Descriptive statistics and narrative analysis are provided followed by the pooling of data and statistical meta-analysis where possible.

3.15.1 Overall measures of 'proving' reactions

One approach to the design of HPTs is to pre define a proving reaction and then to use the number of participants who display this proving reaction as an outcome measure. 3 of the included studies measured this (see table 3.10)

Table 3.10 "proving" reactions in individual trials.

Study	Number of participants demonstrating a proving reaction	Definition of proving reaction
Vickers Van Haselen	4/52 (8%) verum vs. 1/52 (2%) placebo	2 true symptoms on at least 2 consecutive days. A difference score of more than 10 between trial period and baseline on this measure
Goodyear	5/20 (25%) verum 1/27 (3.%)	2 true symptoms on at least 2 consecutive days and no more than one false symptom during the 14 days of the study period.

Brien	14 /101(13.9%) verum 15/105 (14.3%) placebo	2 true symptoms on at least 2 consecutive days and no more than one false symptom during the 21 days of the study period.

Proving reactions are usually defined in relation to the occurrence of a particular number or pattern of symptoms which would be expected as pathogenetic effects from the specific test medicine. For Goodyear (1998) the proportion of individuals experiencing a proving reaction was their secondary outcome and was defined as experiencing at least 2 true symptoms on at least 2 consecutive days and no more than one false symptom during the 14 days of the study period. On this measure 5 from 20 participants in the Belladonna group proved, and 1 from 27 in the placebo group. 1 of the participants in the Belladonna group withdrew but had already provided data as the reason for the withdrawal was a strong proving reaction with 2 true symptoms and no false symptoms. It is unclear why the authors do not include this result in their final analysis.Vickers and Van Haselen (2001), for whom this was also the secondary outcome measure, found no significant differences between groups for the number of participants meeting predefined criteria for a proving reaction. They report two different figures in relation to proving reactions: the number of episodes of true symptoms and then the number of responders (participants experiencing a proving reaction). There were more episodes of true symptoms in the placebo group but, according to their own definition (a difference score of more than 10 or more during the trial period) there were actually 4 provers from 52 in the verum group and 1 from 52 in the placebo group. Brien (2003) reported no significant group differences in proving rates. Rates in this trial were over 14/101 and 15/ 105 : proving rates of 14.3% and 13.9 respectively. were reported, a mean difference of 0.4% (-9.3,10.1)

The studies by Brien and Goodyear both used very similar definitions of a proving reaction, and yet found very different outcomes. Whilst the sample size was much smaller Vickers and Van Haselen found proving rates four times higher in the verum

group. The proving rates in the Goodyear trial are also higher in the verum group, whilst Brien found no difference between verum and placebo groups. A closer look at the definitions of proving reactions suggests a possible explanation for some of this heterogeneity. It is not reported who defined the symptoms which were true and false for the medicine used in the trial, or what process was used to select these symptoms. In both the Brien and the Goodyear trials the medicine was Belladonna 30C, yet looking at the true and false symptoms used to define proving reactions in the two trials there is not a single symptom which is common to both.

Whilst there was therefore some clear heterogeneity between the three studies from which data relating to proving reactions were pooled, a statistical assessment of this heterogeneity gave a value of 55% which is considered moderate. The trials of Brien, Goodyear and Vickers and Van Haselen were measuring similar things with similar enough outcomes measures that it was therefore deemed appropriate to pool the data. As heterogeneity was moderate analysis using both fixed and random effects models was carried out and the results compared.

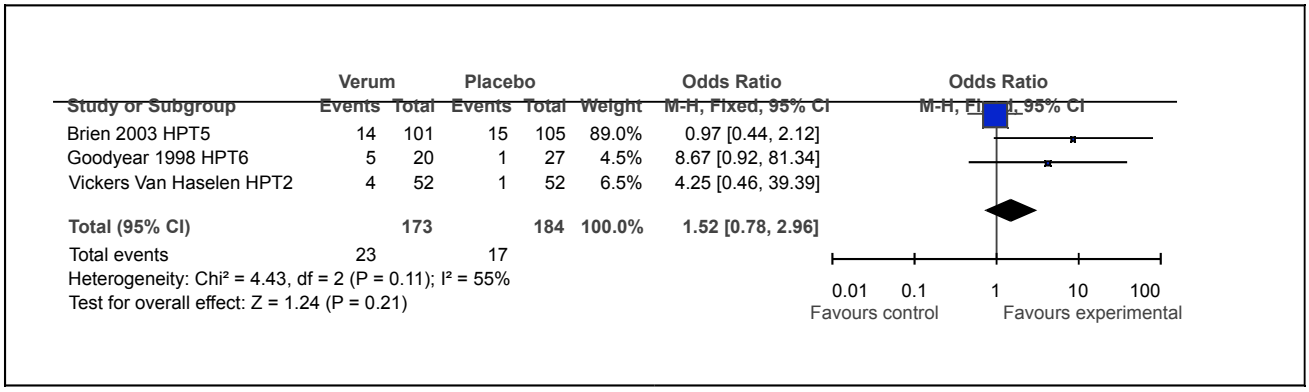


Figure 3.11 Meta analysis (fixed effects model) Participants showing a proving reaction.

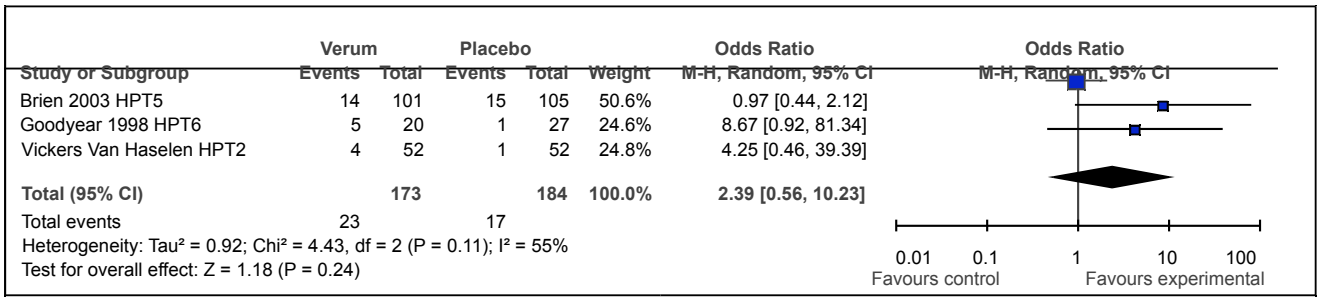


Figure 3.12 Meta analysis (random effects model) Participants showing a proving reaction.

In summary, inconclusive findings were identified from the three trials of varying sizes that looked at overall proving reactions. It is likely that this heterogeneity is related to the variations in outcome measures and to the idiosyncratic reactions of individuals to homeopathic medicines. Pooling of data and meta-analysis demonstrated no significant differences. Hence it is difficult to conclude from this measure whether there are any differences between homeopathic medicines and placebo when given to healthy participants.

3.15.2 Symptoms typical of the test medicine

In total, available data from 9 of the included studies indicates that 2093 symptoms typical of the test medicine occurred in 242 participants taking verum compared to 1353 symptoms in 258 participants taking placebo.

Standardised Mean Differences, where reported in the following paragraphs, have been calculated for this review using mean difference data provided in the reports of the trials. In using this outcome measure authors are assessing differences between those receiving placebo and those receiving homeopathic medicine in the experience of symptoms which might be expected from that medicine.Walach (2001) looked for 9 categories of symptoms which they deemed typical for the test medicine – Belladonna. The only type of symptoms observed more frequently with Belladonna than with either baseline or placebo were throat symptoms (P=0.07). The authors note that the few significant differences found across the different categories were weak and were lost when corrected for the number of tests

performed. Overall the SMD for the groups was 0.13 (CI -0.16,0.43). Vickers and Van Haselen (2001) defined 5 symptoms as true of their test medicine and 5 as false and developed a score for the difference between the two based on the number of days on which true and false symptoms occurred. They subtracted the difference score for the baseline period from the score for the trial period to arrive at their primary outcome measure. There were no significant differences between groups on this measure [SMD 0.03 CI -0.36,0.41] . Goodyear (1998) had, as a primary outcome, the difference in the number of true and false symptoms experienced by those in the verum and placebo groups. The test medicine was again Belladonna in this trial. No significant difference was found between the two groups (the authors state that confidence intervals cannot be calculated because the distributions are non normal). Walach (2004) report no significant differences between symptoms typical and atypical of their test medicine, Cantharis [SMD 0.34,CI -0.8,1.48].

Signorini (2005) had, as a secondary outcome measure the number of symptoms which were concordant with symptoms reported in a previous HPT of the same medicine. They report that, from the number of new or unusual symptoms recorded in their trial of two different medicines and placebo, 30 symptoms from the group taking the first medicine were concordant with a previous HPT of the same medicine (45% of the total) , compared to 10% of the symptoms recorded by those taking a different medicine

Walach (2008) reports two trials, the first of which compared Ozone to placebo and the second of which compared both Ozone and Iridium to placebo. In the first trial the average number of symptoms deemed typical for Ozone was 75 in the verum group and 45 in the placebo group. However, no significant differences were reported between verum and placebo groups in terms of symptoms deemed typical of the test medicine [SMD 0.52, CI -0.47,1.51] In the second trial a highly significant difference was found between verum and placebo groups in terms of symptoms typical for Ozone [SMD 1.17 CI 0.29,2.05]. The difference between Iridium and placebo was not significant [SMD -0.23 CI -1.02,0.55]. Mollinger (2009) reported significant differences between symptoms typical of the two trial medicines across the verum and placebo groups. The two verum groups experienced 5 and 6 symptoms deemed typical of the respective medicines, compared to almost zero

typical symptoms in the placebo group. The data necessary to calculate SMDs for this trial has not been available.

Overall, three of the individual trials reported significant differences – Signorini (2005) Walach (2008) and Mollinger (2009). Unfortunately data was not available from two of these which was suitable for pooling

A lack of consensus in relation to which symptoms can be deemed typical or true of a particular homeopathic medicine means that there may be clinical heterogeneity relating to this outcome measure.

However, data was pooled where possible for this outcome – (symptoms deemed typical of the test medicine) and assessed statistically.Statistical heterogeneity was actually relatively low (I^2 =29%) and a fixed effects model was therefore used to compare outcomes.

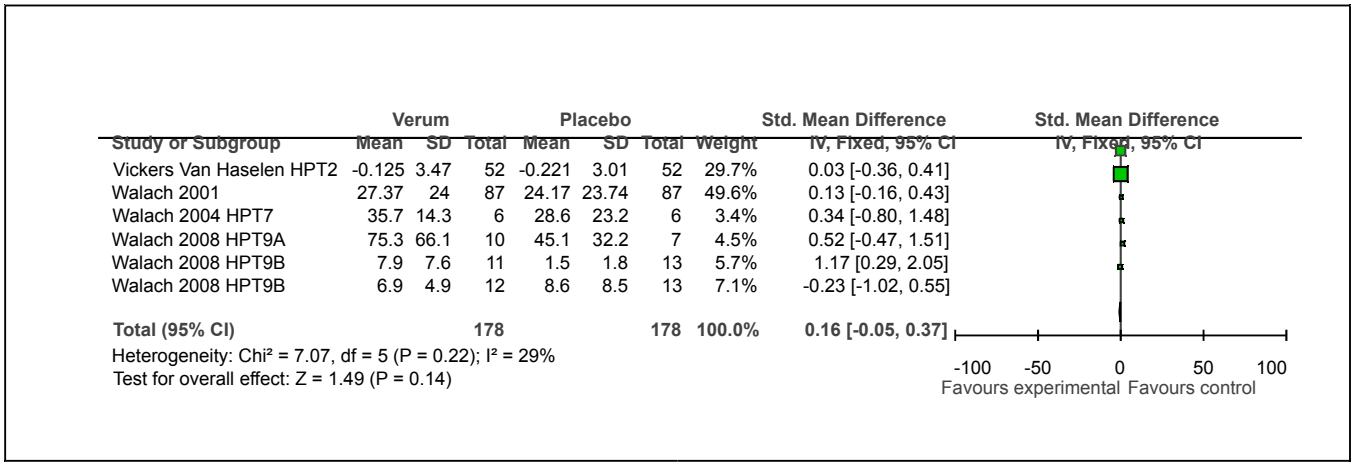


Figure 3.13 Meta-analysis Symptoms typical of the test medicine

Only one of the small trials for which suitable data was available had reported a significant result on this dimension and pooling of results clearly showed that there was no significant difference between the verum and placebo groups on this measure.

3.15.3 Differences between baseline and intervention phases.

Most HPTs measure comparative outcomes across verum and placebo groups; many also involve a baseline period. 7 of the reviewed trials measured differences between baseline and intervention periods in relation to the numbers and pattern of

symptoms occurring. There were two distinct approaches to this, with some trials providing placebo to all participants in this period and some using it as a pure observation period. As discussed in chapter one, it is known that the number and pattern of symptoms produced by placebo interventions and the number of background symptoms reported on a daily basis by healthy volunteers both vary widely but can be large. Comparing the change in symptom pattern from a baseline period to a period of either verum or placebo intervention is therefore instructive.

Table 3.14 Differences in symptoms reported between baseline run in phases and intervention phases of trials.

Trial and nature of baseline run in period	Measures Used	Outcomes
Walach(2001) Observation only	No of symptoms typical of medicine. Differences between baseline and trial for both placebo and verum groups	No significant difference Baseline run in – 24.26(22.15) Placebo:24.17(23.74) Verum : 27.37(24.00)
Walach (2004) Observation only	No of symptoms typical of medicine. Differences between baseline and trial for both placebo and verum groups	Significant difference between baseline and experimental period for all symptoms (p=0.03 for typical sx and p=0.02 for atypical sx)
Signorini Observation only	Qualitative comparisons between baseline (observation only) period and trial period	Qualitative measures only
Walach (2008) Observation only	No of symptoms typical of medicine. Differences between baseline and trial for both placebo and verum groups	Significant difference. Baseline run in 21.1 (SD20.1) Placebo: 45.1(SD32.2) Verum: 75.3(SD66.1)
Vickers	Symptom score in placebo	No significant differences-

Placebo run in	run in phase subtracted from symptom score in trial phase.	statistics for differences between run in and experimental phase not reported
Brien Placebo run in	Proving reactions. Differences between baseline and trial for both placebo and verum groups	Placebo run in – 7.9% and 6.7% Treatment phase – 13.9% and 14.3%

Walach (2001) noted that theoretically one might expect the lowest frequency of symptoms during baseline with no intervention, followed by placebo and then verum. In their trial instead they frequently found unexpected patterns with certain symptoms higher in the baseline period than in either placebo or verum groups. In that trial there were no significant differences in relation to symptoms deemed typical for the trial medicine between baseline, placebo and verum results. Walach in a later trial (2004) noted a significant increase in both symptoms deemed typical of the trial medicine and in atypical symptoms between the one week baseline and the two week intervention period, for both placebo and verum and groups. Signorini (2005) used a one week run in period as an observational period. Qualitative comparisons were made between the diary reports of patterns of symptoms in the intervention and the baseline periods. In two trials reported by Walach (2008) there were obvious differences between baseline and intervention. In the first trial the mean number of symptoms deemed typical of the trial medicine was 21.1 (SD20.1) in the baseline period, and 45.1(SD32.2) and 75.3(66.1) in the placebo and verum groups during the intervention phase. These figures do not reach statistical significance. In the second trial, two medicines were compared to placebo. The mean number of symptoms typical for the first medicine (Ozone) was 0.9 (1.6) at baseline, and in the trial was 1.5(1.8) in the placebo group, 3.6(2.8) in the Iridium group and 7.9(7.6) in the Ozone group. For the group taking the second medicine, the pattern is less clear. In fact one of the statistically significant results that the authors note is the counter-intuitive

finding that symptoms deemed typical of Iridium are higher in the baseline phase than in the experimental phase (p=0.037)

Vickers (2001) used a one week placebo run in and symptom scores during this period were subtracted from symptom scores during the intervention phase to arrive at the primary outcome – a score to determine whether a proving reaction had taken place. Brien (2003) also used a one week placebo run in. They found no difference in proving rates between the verum and the placebo group. They do not comment on the differences between the run in phase (in which both groups received placebo) and the intervention phase but the data clearly shows that proving rates were twice as high in the intervention phase for both groups. (13.9 and 14.3% compared to 7.9% and 6.7%)

Where the data was available there were significant differences between the run in period and the experimental period in 3 of the 5 trials. One of these involved a one week placebo run in; the others were observation only periods. Such results suggest that the symptoms which occur during an intervention phase (both in placebo and verum groups) are different from those which occur in run in periods. Such an outcome is consistent with the idea that expectancy effects can be significant: when people expect that they may receive an intervention they are more likely to display symptoms. It could also be argued that the observation run in period could act as a learning period in which people progressively become more aware of symptoms and of what is required in terms of observing and reporting. If this were the case then differences between baseline and intervention would be an artefact of learning through time rather than a real difference relating to the intervention.

3.15.4 Symptoms in typical categories

As noted earlier, expert guidance on the conduct of HPTs has suggested that data from such trials should be gathered in particular categories. Only one of the papers reported measures of the numbers and proportions of symptoms in each of these categories. (Dominici 2006). Using data from two small trials they found that new symptoms constituted 174 (46%)and 83 (44%) in the verum groups against 35(34%)

in the placebo group. The figures for exceptional symptoms were 49 (13%) and 29 (15.5%) against 2(1%). Cured symptoms were 11(3%) and 9 (4.5%) against 7%. (8%).The return of old symptoms were 75(20%) and 36 (19%) against 8 (7%). Common symptoms provided 68(18%) and 32(17%) of the symptoms in the verum groups and 55(51%) of those in the placebo group. The authors report chi square bivariate analysis of the number of symptoms in each category in the three groups. Differences were found at high levels of confidence (P <0.001). The same control group was used as a comparator for the two separate trials reported.

Table 3.15 Proportions of symptoms occurring in different categories

Category	Dominici 2006 HPT3A	Dominici 2006 HPT3B
New symptoms	V = 174 (46%) C= 35(34%) V = verum C = control	V= 83(44%) C= 35 (34%)
Old symptoms	V= 75(20%) C= 8 (7%)	V= 36(19%) C= 8(7%)
Cured	V = 11 (3%) C = 7 (8%)	V= 9 (4.5%) C = 7(8%)
Exceptional	V = 49(13%) C = 2 (1%)	V = 29(15%) C = 2 (1%)
Common	V = 68(18%) C = 55 (51%)	V = 32 (17%) C = 55 (51%)

Signorini (2005) measured the number of phrases recorded by participants in diaries which contained unusual or new symptoms and the return of old symptoms. However, numerical data relating to these outcomes are not reported.

Given the small number of participants in the Dominici trials (n=21) and the lack of any other trials reporting data on these outcomes, no conclusions can be drawn in relation to this outcome.

3.15.5 Placebo/Nocebo.

Means and percentages for the individual trials are reported below.

Table 3.16 Mean no of symptoms in placebo groups and percentage of participants in placebo groups reporting pathogenetic symptoms.

Trial	**Mean no of symptoms reported in placebo groups.**	**Trial**	**Percentage of participants in placebo group reporting pathogenetic symptoms**
Walach (2001)	24.17	Signorini (2005)	25-30% - any pathogenetic symptom
Dominici (2005)	18	Fisher (2001) Fisher (2001)	70% - symptoms compatible with the trial medicine 40%- symptoms compatible with the trial medicine
Walach (2004)	28.6 16.2	Vickers, van Haselen (2001)	40% - 2 symptoms on 2 consecutive days
Signorini	4.5 (new or unusual	Brien (2003)	14.3% - symptoms compatible

(2005)	symptoms only)		with the trial medicine
Walach (2008)	64 12	Goodyear (1998)	66% - at least one symptom
Mollinger(2009)	11		

Walach (2001) found a mean of 24.17 symptoms in participants receiving placebo during the trial. In the trials by Dominici et al the mean was 18. Walach (2004) reported mean numbers of symptoms typical and untypical for the trial medicine in the placebo group as 28.6 and 16.2. Signorini reported that 58 phrases containing new or unusual symptoms were recorded by the placebo group – a mean of 4.5 per person. Walach (2008) reported two trials, with mean numbers of symptoms in the placebo groups at 64 and 12 respectively. Mollinger (2009) reported a mean of 11 symptoms in the placebo group.

Signorini and colleagues (2005) pointed to several distinct responses to placebo intervention. They noted that 25-30% of participants developed apparent pathogenetic symptoms and that this nocebo effect is similar to that described elsewhere. Fisher (2001) in two trials recorded how many participants in placebo groups reported symptoms that might be deemed compatible with the trial medicine on a pathogenetic index. Rates were 70% and 40%. Vickers reported that 40% of placebo participants reported symptoms that had been deemed true for the test medicine.Brien (2003) reported proving reactions in 14.3% of the placebo group. Such a reaction was defined as at least two true symptoms on consecutive days and no more than one false symptom. Goodyear reported that 66% (18 from 27) participants in the placebo group recorded at least one symptom.

The discussion in chapter one highlighted the impact of expectancy and Hawthorne effects on the reporting of symptoms. The results found in this review demonstrate a very high level of placebo or nocebo effects. For the narrower set of symptoms which may be classed as typical of the test medicine the percentage of participants in placebo groups displaying these symptoms is very high.(up to 70% and otherwise between 25 and 40%). Those displaying any symptoms were up to 66%. Some of these symptoms may have been pre-existing symptoms since the definition of healthy participant at entry was often loosely defined and did not preclude the existence of some symptoms.

CHAPTER FOUR – DISCUSSION

4.1 Summary of findings

Meta –analysis of pooled data found no significant quantitative difference between homeopathic medicines and placebo medicines in terms of pathogenetic effects.

4.2 Strengths and weaknesses of this review.

This study is the first to review HPT's which meet modern standards for trial design and methodological quality. Including only such studies means that the results are likely to be less biased. However, in terms of homeopathic theory all of the studies shared a weakness, in that none of them selected volunteers for sensitivity to the medicine. Research into homeopathy must be robust and of high quality in terms of adopting accepted standards of trial design. It must also adopt designs which are appropriate to the intervention. All of the homeopathic literature suggests that only a minority of individuals will respond to a particular homeopathic medicine. This is true both in clinical practice and also in pathogenetic trials. The exact proportion of individuals who will respond to a given medicine is unknown but from data included in this review the level of proving reactions and pathogenetic effects is unlikely to exceed 14% of any random sample of the population. Levels of nocebo effects in

any random sample of the population are likely to be at least at the same level. Therefore HPT's which involve random samples of the population are unlikely to show quantitative differences between placebo and intervention groups.

Several aspects of methodological quality can have a direct bearing on outcomes. Reviews often use sensitivity analyses to re-assess statistical outcomes according to the inclusion or exclusion of different studies. (Crowther and Cook 2007).For example excluding studies which are of lower size, or which did not have clear procedures for randomisation. Such analyses are also used to compare the effects of imputing missing data. In this review there was insufficient data to perform such sensitivity analyses. Most of the studies were of good quality in relation to the key design features as discussed. However, the included trials varied in size and it is known that trial size can be a key factor in terms of the ability to detect true outcomes.Trials with insufficient participants will be underpowered to detect real differences in outcomes (type II errors) for example (Freiman et al 1978)

9 of the 15 trials included in this review had less than 50 participants.

It is highly likely that language bias affects the outcomes of this review.The study was limited to studies for which a full translation was available in English. It is known that many studies in the field of complementary medicine are published in languages other than English. Some 15 % of the papers identified by the search strategy were not fully available in English and were excluded from this review. Whilst the effect of including or excluding non English papers in systematic reviews is not fully clear studies have shown that, for conventional medicine, non English language papers generally show smaller effect sizes, but that non English papers relating to complementary medicine show the opposite, with larger effect sizes. (Bartlett 2000).

Publication bias is also likely to be a factor in this review. It is known that studies with results that are significant, studies of high quality and those that are large and/or well-funded are more likely to be submitted and accepted for publication, than studies without such characteristics. (Song et al 2000). A recently published Cochrane review noted that trials with positive findings (defined either as those that were statistically significant ($P < 0.05$), or those findings perceived to be important or striking, or those indicating a positive direction of treatment effect), had nearly four

times the odds of being published compared to findings that were not statistically significant (P ≥ 0.05), or perceived as unimportant, or showing a negative or null direction of treatment effect. Whilst publication bias is a real problem and it may be present in up to 50% of meta-analyses (Sutton et al 2000) in most cases it does not seem to affect the conclusions of the reviews concerned (Sutton et al 2000)

One way in which publication bias is widely assessed is by the use of the funnel plot. (Cullinan 2005).This is a graphical display which plots the effect size reported in each individual trial against some measure of the size of the trial, such as the overall sample size or the standard error. The plot is so named because the shape of the plot should appear like a funnel if no publication bias exists. This is because trials of smaller size, which are always more numerous, show a large variation in the estimates of effect size that they report. This is because random variation has a great deal more influence in small trials. At the narrower end of the funnel fewer trials of larger size will show a narrower range of variation in effect size estimates.

Funnel plots are only useful where there is an adequate distribution of studies with varying sample sizes. It is also an unwarranted assumption to conclude that asymmetry in a funnel plot automatically means that there is publication since other factors can also lead to such asymmetry, such as differences in the quality of similar sized studies, clinical heterogeneity and chance. Given these limitations and the small number of studies available for this review, it was deemed that a funnel plot would not add any useful information to this review.

4.3 Meaning of this review.

The outcomes are consistent with the hypothesis that homeopathic medicines do not produce pathogenetic effects in healthy volunteers. An alternative explanation for the findings would be relate to the hypothesis outlined in chapter one that pathogenetic effects will only occur in a minority of sensitive volunteers. Goodyear (1998) started with the hypothesis that 10 to 20% of participants were likely to exhibit proving reactions in HPT's. Walach had previously reported proving rates of just 1% (Walach 1994).Data analysed from the studies included in this review found proving rates of 13% in verum groups and 9% in placebo groups. This hypothesis has been part of homeopathic theory since its inception. Hahnemann, who first developed

homeopathy suggested this in his book the Organon of Medicine which has continued to be the most influential publication in homeopathic medicine since it was first published in1810. He stated that strong medicines will bring about alterations in health, even in robust people. He then suggested that weaker ones , on the other hand, " *reveal their true action only when tested on delicate susceptible and sensitive people who are free from disease*". (Hahnemann 1921)

It is somewhat puzzling then that none of the trials which were otherwise of high quality and met the criteria for inclusion in this review, used a design which allowed for screening and selection of sensitive volunteers. Such a design could involve a two stage process. Stage one of the trial could be conducted with the aim of determining which participants are sensitive to the particular medicine being tested. Careful quantitative and qualitative analysis of apparent pathogenetic effects in stage one would be carried out by persons not involved in the trial. Judgements could then be made about which participants have displayed sensitivity to the medicine and real pathogenetic effects. In stage two this group would be randomised to receive the same medicine again or placebo. Participants and supervisors would remain blinded throughout both stages. A major disadvantage of this proposal is the kind of sample size that might be required at the outset to produce enough participants of the required sensitivity. If just 13% of participants demonstrate proving reactions, then a sample almost eight times larger would be needed for stage one in order to produce enough sensitive volunteers for a sample with enough power for stage two.

However, a small group in stage two who have been predetermined as sensitive to verum are more likely to display effects that are significantly different from placebo than a large group of participants who have been recruited randomly in relation to sensitivity.

Studies described in the Background section reported rates of nocebo symptoms of up to 40%. Studies included in this review found similar rates for a smaller set of only those symptoms which might be linked to the medicine on trial, suggesting that rates for nocebo symptoms overall would be higher. It was noted in chapter one that nocebo symptoms are related to expectations and also to a range of other factors including gender. In HPTs the purpose is to generate pathogenetic symptoms and so

a high level of expectation is created in participants. The majority of participants in HPTs are female – in the studies included in this review the level was 68.5%. Both of these features – high expectations and female gender – are likely to lead to a higher level of nocebo symptoms being reported in these trials than in other trials.

Data from these studies do suggest that overall rates of nocebo symptoms in the populations, settings and contexts which are typical of HPTs may be as high as 70%. Nocebo symptoms which are similar to the symptoms expected from a homeopathic medicine seem to occur in 40% of participants, and nocebo symptoms indistinguishable from a proving reaction in 14%.

If only one in eight participants typically show a proving reaction, then this is unlikely to be distinguishable from the high levels of nocebo symptoms which occur in HPTs.

This data also suggests that HPT's without placebo comparator groups and careful selection of symptoms will report high percentages of symptoms which are caused by factors other than the pathogenetic effects of the specific medicine. Homeopaths do often use careful qualitative analysis before using symptoms from HPT's to guide practice but it remains likely that many symptoms which are recorded in the literature against specific medicines remain unreliable, since many of these attributions originate from poor quality HPT's without adequate controls.

4.4 Unanswered questions and suggestions for future research

It remains unclear whether homeopathic medicines can produce pathogenetic symptoms in healthy volunteers which are quantitatively different from placebo effects.

Only a small fraction of studies which were located met the inclusion criteria – some 3%. This concurs with the findings of the review by Dantas et al (2005) of the quality of HPT's, which suggested that quality had improved in recent years, but remained low in many trials. Those who plan and design HPT's should follow both the extensive guidance which is available in relation to RCT's and that which is available in relation to HPT's.

The most important aspects of these , in terms of reducing bias, and in terms of increasing the likelihood of distinguishing genuine pathogenetic effects from placebo effects and background noise would be:

- Randomisation, using explicit procedures
- Use of placebo comparator groups
- Blinding of participants and researchers and verification of blinding
- Dealing with loss to follow up and adverse events appropriately.
- Using validated outcomes measures
- Operationalising the definition of health which is used as a standard inclusion criteria for HPTs. The use of the SF36 as Fisher used, or some other well validated and widely used measure is suggested.

In order to determine whether homeopathic medicines can produce symptoms which are quantitatively different from placebo in healthy volunteers it is also suggested that future HPTs adopt some kind of framework for selecting sensitive volunteers.In relation to outcome measures, counting symptoms which have been predefined as linked to the test medicine seems to be a frequently used strategy. It has been noted that there are problems in relation to verification and agreement of such symptoms. Two of the included studies used homeopathic belladonna but there was no overlap at all between the sets of symptoms which had been chosen as typical for belladonna. This example highlights the need to validate and reach consensus on such measures. It is suggested that a panel of materia medica experts reach agreements.

Vickers and Van Haselen concluded at the end of their study that closed questionnaires listing a limited number of symptoms are not a useful mechanism for investigating drug proving phenomenon because such phenomena are rare and idiosyncratic. This is consistent with the ideas discussed in chapter one.It could be added that such methods also rely on accuracy and agreement in selecting the closed list of symptoms, and that these have not been demonstrated in HPTs so far.

If such verification of specific symptoms is problematic a more useful measure of outcomes may be to compare symptoms which occur in particular categories which are expected according to the theory of HPTs. This approach is recommended in much of the guidance relating to HPTs (e.g. Wieland 1997, Riley 1997, ECH 1995) but was taken in only one of the papers included in this review (Dominici 2006). The results of the two small trials included in this paper were significant and it is suggested that larger HPTs are conducted using similar outcome measures i.e. those that assess symptoms in the generally categories which are utilised in homeopathic practice to assess response to treatment.

In terms of general guidance the REDHOT guidelines (Reporting Data on homeopathic Treatments (RedHot): A supplement to CONSORT, Dean et al 2006)which were specifically developed to improve the conduct and reporting of trials in homeopathy should be updated to include guidance specifically relating to the design, conduct and reporting of HPTs.

Given the widespread use of homeopathy and the uncertainty that remains in relation to HPTs there is a clear need to develop appropriate research methodologies for HPT's, to plan and carry out HPTs which can determine whether homeopathic medicines can be distinguished from placebo in pathogenetic trials, and to conduct HPT's which can verify that information which has been used in practice from unreliable HPTs.

Altman DG J Bland M Statistics notes (1995) Absence of evidence is not evidence of absence *BMJ* 1995;311:485 (19 August)

Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. BMJ. 1996;313:570-1

Bellavite P and Signorini A (1995) The Emerging Science of Homeopathy Complexity Biodynamics and Nanopharmacology North Atlantic Books Berkeley

Bellavite P Chirumbol s Signorini A and Bianchi A (1991) *Effects of various homeopathic drugs on superoxide production and adhesion of human neutrophils* 5th giri meeting Paris Abs 2

Bellows HP.(1906) The Test Drug-Proving of the O. O. & L. Society : A Reproving of Belladonna Being an Experimental Study of the Pathogenic Action of that Drug Upon the healthy Human Organism. Boston: The O. O. & L. Society,

Belon P (1995) For Debate: Provings Concept and Methodology *British Homeopathic Journal* ,84,pp 213-217

Berlin JA (1997). Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet 350:185-186.

Bildet, J Guyot, M . Bonini, F et al., 'Demonstrating the Effects of Apis mellifica and Apium virus Dilutions on Erythema Induced by U.V. Radiation on Guinea Pigs', *Berlin Journal of Research in Homeopathy***,** 1990, 1:28.

Boissel J: Critical Literature Review on the effectiveness of Homeopathy: Overview over data from homeopathic medicine trials. In: Commission of the European communities, editor. Homeopathic Medicine Research Group. Commission of the European communities; 1996, 196-210.

Bornhöft et al *(2006).* Effectiveness, Safety and Cost-Effectiveness of Homeopathy in General Practice –Summarized Health Technology Assessment*. Forsch Komplementärmed 2006;13(suppl 2):19-29.*

Brien S Lewith G and Bryant T Ultramolecular homeopathy has no observable clinical effects. A randomized, double-blind, placebo-controlled proving trial of Belladonna 30C *Br J Clin Pharmacol* 56, 562–568

Brien S and Lewith G 2005) Assessing Homeopathic Proving Using Questionnaire Methodology: Consideration and Implications for Future Studies *Forsch Komplementärmed Klass Naturheilkd 2005;12:152–158*

Bowling A and Ebrahim S (2005) *Handbook of Health Research Methods* Open University Press Maidenhead

Carstens Stiffung (2008) HomBRex - a database on Basic Research experiments on Homeopathy www.carstens-stiftung.org/ [accessed on 3/7/09]

Calderon MA, Alves B, Jacobson M, Hurwitz B, Sheikh A, Durham S (2007). Allergen injection immunotherapy for seasonal allergic rhinitis. *Cochrane Database of Systematic Reviews* 2007, Issue 1. Art. No: CD001936

Conforti A Signorini A Bellavite P (1993) Effects of high dilutions of histamine and other natural compounds on acute inflammation in rats in *Omeomed 92 Editrice Compositori Bologna p163*

Cucherat, M., Haugh, M. C., Gooch, M., & Boissel, J. P. (2000), Evidence of clinical efficacy of homeopathy. A meta-analysis of clinical trials. HMRAG. Homeopathic Medicines Research Advisory Group", *Eur.J.Clin.Pharmacol., vol.. 56, no. 1, pp. 27-33.*

Cullinan P (2005) *Evidence based Health Care Systematic Reviews in* Bowling A and Ebrahim S (2005) Handbook of Health Research Methods Open University Press Maidenhead

Dantas F, Fisher P, Walach H et al. (2007) A systematic review of the quality of homeopathic pathogenetic trials published form 1945 10 1955. *Homp* 2007; 96

Dantas, F. & Rampes, H. (2000). Do Homeopathic Medicines Provoke Adverse Effects? A Systematic Review. *British Homeopathic Journal. 2000; 89: 70-74.*

Deans M E (2004) *The Trials of Homeopathy* KVC Verlag Germany

De Gerlache A and Lans M (1991) Modulation of experimental rat liver carcinogenesis byultra low doses of the carcinogens in *Ultra Low Doses* (C Doutrempuich ed) Taylor and Francis London

De Marque D (1987) the Developmetn of Proving Methods Since Hahnemann *British Homeopathic Journal*, 76 ,pp71-75

Egger M, Smith GD, Altman DG (2001) Systematic reviews in Health Care: Meta analysis in context. London, BMJ Publications.

Edwards, P., M. Clarke, C. DiGuiseppi, S. Pratap, I. Roberts, and R. Wentz. 2002. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine 21: 1635-1640.*

European Committee for homeopathy (2004) Homeopathic drug proving guidelines European Committee for homeopathy Brussels.

European Council for Classical Homeopathy (2007) ECCH Guidelines for Homeopathic Pathogenic Trials Second Edition ECCH Norfolk UK

Faculty of Homeopathy (2008) Systematic Reviews of RCT's in Homeopathy http://www.facultyofhomeopathy.org/research/systematic_reviews/ [Accessed 7/5/09]

Fimiani V Cavallaro A Ainis T and Bottari C (2000) Immunomodulatory Effect of the Homoeopathic Drug Engystol-N on Some Activities of Isolated Human Leukocytes and in Whole Blood *Immunopharmacology and Immunotoxicology* 2000, Vol. 22, No. 1, Pages 103-115

Global TGI (2008) Global TGI Barometer, January 2008; Issue 33.

Heirs M and Dean ME Homeopathy for attention deficit/hyperactivity disorder or hyperkinetic disorder (review) *Cochrane Database of Systematic Reviews 2007 Issue 4 Art No CD005648*

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008

Hill C, Doyon F. (1990)Review of randomized trials of homeopathy. *Rev Epid'em et Sant'e Publ 1990; 39: 139-147*

Jonas W, Lyn Y, Tortella F (2001) Neuroprotection from glutamate toxicity with ultra low dose glutamate *Neuroreport 12 (2) 335*

Kaptchuk T J (1996) When Does Biased Become Unbiased The Dilemma of Homeopathic Provings and Modern Research Methods *British homeopathic Journal October 1996 85,237-247*

Kienle, G. S. & Kiene, H. (2001) A critical reanalysis of the concept, magnitude and existence of placebo effects. In *Understanding the Placebo Effect in Complementary Medicine. Theory, Practice and Research* (ed. D. Peters), pp. 31–50 London: Churchill Livingstone.

Kleijnen J et al: Clinical trials of homoeopathy. BMJ 1991;302:960 (20 April)

Koster d, Van Haselen R, Jansen G and Dicke M (1998) Homeopathic medicine test according to the crossover design:from theory to practice. *British Homeopathic Journal 87 (4) pp181-189*

Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, Jonas WB (1997). Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet. 1997 Sep 20;350(9081):834-43.*

Mathie, R (2003). The research evidence base for homeopathy: a fresh assessment of the literature. *Homeopathy 92: 84-91. 2003.*

McCarney RW, Linde K, Lasserson TJ. Homeopathy for chronic asthma. (2004) *Cochrane Database of Systematic Reviews* 2004, Issue 1. Art. No.: CD000353.

Moher, D Schulz K, AltmanD G,(2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials *The Lancet 357 April 14, 2001*

Moher D, Horsley T, Morgan K, Tricco A, Tsertsvadze A (2007) Blinded versus unblinded assessments of risk-of-bias in studies included in a systematic review (Protocol). *Cochrane Database of Systematic Reviews* 2007, Issue 4. Art. No.: MR000025.

Peters D (Ed) (2001) *Understanding the Placebo Effect in Complementary Medicine. Theory, Practice and Research* Churchill Livingstone London

Riley D S (1997) Extracting Symptoms from homeopathic drug provings *British Homeopathic Journal* 86 ,225-228

Schulz KF, Chalmers I, Hayes RJ, and Altman DG.(1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273 408-12

Silagy C Middleton P and Hopewell S (2002) Publishing protocols of systematic reviews comparing what was one with what was planned *JAMA* 287 (21) 2831-4

Shalts E (2002) A Double Blind Placebo Controlled Pilot Study of the Validity of Provings as a method of Finding Pathogenic Characteristics of Homeopathic Remedies. In Herscu P Provings Volume 2 New England School of Homeopathy Press Amherst MA

Shang A Huwiler-Muntener K , L. Nartey L *et al.*, (2005) Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy, *Lancet* **366** (2005), pp. 726–732.

Sherr J (1994) The dynamics and methodology of Homeopathic Provings Dynamis Books Malvern.

Stolberg M. Die Homo¨ opathie auf dem Pru¨ fstein. Der erste Doppelblindversouch der Medizingeschichte im Jahr 1835. Munchener Medizinische Wochenschrift 1996; 138: 364–366.

Taylor MA, Reilly D, Llewellyn-Jones RH, McSharry C, Aitchison T, Lancaster T, Vickers A.(2000) Randomised controlled trial of homeopathy versus placebo in perennial allergic rhinitis with overview of four trial series *British Medical Journal 2000; 321: 471-476.*

Thompson T (2004). A preliminary audit investigating remedy reactions including adverse events in routine homeopathic practice. *Homeopathy 93;203-209*

Toper et al 1990 Effects of high dilutions of antigens on the generation of specific antibodies in proc 4th GIRI meeting Paris Abs 15

Torgerson DJ and Torgerson CT (2008) *Designing Randomised Trials in Health, Education and the Social Sciences* Macmillan London

Truzzi (1998) On Some Unfair Practices towards Claims of the Paranormal in Binkowski,E (1998) *Oxymoron: Annual Thematic Anthology of the Arts and Sciences, Vol.2:* The Fringe, New York: Oxymoron Media

Wagner H Kreher B Jurcic K In *vitro stimulation* of human granulocytes and lymphocytes by pico- and femtogram quantities of cytostatic agents *Arzneimittleforschung Drug Res 38 273*

Weatherley Jones E Thompson EA and Thomas KJ (2004) The placebo controlled trial as a test of complementary and alternative medicine:observations from research experience of individualised homeopathic treatment *Homeopathy 93 186-189*

Wieland F (1997) The need for GHP guidelines. A brief survey of recent developments in methodology of homeopathic drug provings in Europe *Brit Hom J* 86 229-234

Wieland F (1998) The role of drug proving in the homeopathic concept in Enrst E and Eckhart EG *Homoeopathy a Critical Appraisal*. Butterworth – Heinemann Oxford

Witt CM, Bluth M, Albrecht H, et al. The in vitro evidence for an effect of high homeopathic potencies – a systematic review of the literature. *Complementary Therapies in Medicine, 2007; 15: 128–138.*